

普通高等教育“十一五”国家级规划教材

“信息化与信息社会”系列丛书之

高等学校信息管理与信息系统专业系列教材

信息与知识管理

王众托 吴江宁 郭崇慧 编 著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书是一本以知识管理为主要内容的教材,由于知识管理与信息管理有着密切的联系,因此书中也对信息管理做了概括介绍,但主要讲述知识管理内容。书中首先简要地介绍了信息与信息管理,以及知识与知识管理的基本概念。对信息管理与知识管理具有共性和关联的一些方法和工具,合并在一起进行介绍。本书着重阐述知识管理系统的结构和工作过程,以及知识管理的应用和知识系统的开发。

本书既可作为高等学校信息管理与信息系统专业(偏重于信息系统开发与应用)的本科知识管理课程教材,也可作为其他专业本科及专业学位研究生课程和通识课程教材,以及企事业单位培训教材,还可供从事创新与创业的人士阅读参考。教材内容可以灵活加以组织和裁减。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

信息与知识管理 / 王众托, 吴江宁, 郭崇慧编著. —北京: 电子工业出版社, 2010.6
(“信息化与信息社会”系列丛书. 高等学校信息管理与信息系统专业系列教材)
普通高等教育“十一五”国家级规划教材
ISBN 978-7-121-10895-2

I. ①信… II. ①王… ②吴… ③郭… III. ①信息管理—高等学校—教材②知识经济—应用—管理—高等学校—教材 IV. ①G203②C93

中国版本图书馆 CIP 数据核字(2010)第 090228 号

策划编辑: 刘宪兰

责任编辑: 刘宪兰

印 刷:

装 订:

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 21.5 字数: 465 千字

印 次: 2010 年 6 月第 1 次印刷

印 数: 4 000 册 定价: 35.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。



作者简介

王众托，大连理工大学管理学院教授，中国工程院院士，长期从事系统与信息学科的教学、科研工作，以及教材编写的组织工作。主要教学与研究领域包括系统工程、决策方法、企业信息化和知识管理等。

吴江宁，工学博士，大连理工大学管理学院教授，主要从事知识系统建模、文本挖掘、信息检索、知识可视化等方面的教学和科研工作。

郭崇慧，工学博士，大连理工大学管理学院教授、博士生导师，多年来一直从事运筹学与管理科学（OR/MS）、数据挖掘与机器学习、商务智能与知识管理等方面的教学和科研工作。





总 序

信息化是世界经济和社会发展的必然趋势。近年来，在党中央、国务院的高度重视和正确领导下，我国信息化建设取得了积极进展，信息技术对提升工业技术水平、创新产业形态、推动经济社会发展发挥了重要作用。信息技术已成为经济增长的“倍增器”、发展方式的“转换器”、产业升级的“助推器”。

作为国家信息化领导小组的决策咨询机构，国家信息化专家咨询委员会一直在按照党中央、国务院领导同志的要求就信息化前瞻性、全局性和战略性的问题进行调查研究，提出政策建议和咨询意见。在做这些工作的过程中，我们愈发认识到，信息技术和信息化所具有的知识密集的特点，决定了人力资本将成为国家在信息时代的核心竞争力，大量培养符合中国信息化发展需要的人才已成为国家信息化发展的一个紧迫需求，成为我国应对当前严峻经济形势，推动经济发展方式转变，提高在信息时代参与国际竞争比较优势的关键。2006年5月，我国公布《2006—2010年国家信息化发展战略》，提出“提高国民信息技术应用能力，造就信息化人才队伍”是国家信息化推进的重点任务之一，并要求构建以学校教育为基础的信息化人才培养体系。

为了促进上述目标的实现，国家信息化专家咨询委员会一直致力于通过讲座、论坛、出版等各种方式推动信息化知识的宣传、教育和培训。2007年，国家信息化专家咨询委员会联合教育部、原国务院信息化工作办公室成立了“信息化与信息社会”系列丛书编委会，共同推动“信息化与信息社会”系列丛书的组织编写工作。编写该系列丛书的目的，是力图结合我国信息化发展的实际和需求，针对国家信息化人才教育和培养工作，有效梳理信息化的基本概念和知识体系，通过高校教师、信息化专家、学者与政府官员之间的相互交流和借鉴，充实我国信息化实践中的成功案例，进一步完善我国信息化教学的框架体系，提高我国信息化图书的理论和实践水平。毫无疑问，从国家信息化长远发展的角度来看，这是一项带有全局性、前瞻性和基础性的工作，是贯彻落实国家信息化发展战略的一个重要举措，对于推动国家的信息化人才教育和培养工作，加强我国信息化人才队伍的建设具有重要意义。

考虑当前国家信息化人才培养的需求、各个专业和不同教育层次（博士生、硕士生、本科生）的需要，以及教材开发的难度和编写进度时间等问题，“信息化与信息社会”系列丛书编委会采取了集中全国优秀学者和教师、分期分批出版高质量的信息化教育丛书

的方式,根据当前高校专业课程设置情况,先开发“信息管理与信息系统”、“电子商务”、“信息安全”三个本科专业高等学校系列教材,随后再根据我国信息化和高等学校相关专业发展的情况陆续开发其他专业和类别的图书。

对于新编的三套系列教材(以下简称系列教材),我们寄予了很大希望,也提出了基本要求,包括信息化的基本概念一定要准确、清晰,既要符合中国国情,又要与国际接轨;教材内容既要符合本科生课程设置的要求,又要紧跟技术发展的前沿,及时地把新技术、新趋势、新成果反映在教材中;教材还必须体现理论与实践的结合,要注意选取具有中国特色的成功案例和信息技术产品的应用实例,突出案例教学,力求生动活泼,达到帮助学生学以致用目的,等等。

为力争出版一批精品教材,“信息化与信息社会”系列丛书编委会采用了多种手段和措施保证系列教材的质量。首先,在确定每本教材的第一作者的过程中引入了竞争机制,通过广泛征集、自我推荐和网上公示等形式,吸收优秀教师、企业人才和知名专家参与写作;其次,将国家信息化专家咨询委员会有关专家纳入到各个专业编委会中,通过召开研讨会和广泛征求意见等多种方式,吸纳国家信息化一线专家、工作者的意见和建议;最后,要求各专业编委会对教材大纲、内容等进行严格的审核,并对每一本教材配有一至两位审稿专家。

如今,我们很高兴地看到,在教育部和原国务院信息化工作办公室的支持下,通过许多高校教师、专家学者及电子工业出版社的辛勤努力和付出,“信息化与信息社会”系列丛书中的三套系列教材即将陆续和读者见面。

我们衷心期望,系列教材的出版和使用能对我国信息化相应专业领域的教育发展和教学水平的提高有所裨益,对推动我国信息化的人才培养有所贡献。同时,我们也借系列教材开始陆续出版的机会,向所有为系列教材的组织、构思、写作、审核、编辑、出版等做出贡献的专家学者、教师和工作人员表达我们最真诚的谢意!

应该看到,组织高校教师、专家学者、政府官员及出版部门共同合作,编写尚处于发展动态之中的新兴学科的高等学校教材,还是一个初步的尝试。其中,固然有许多的经验可以总结,也难免会出现这样那样的缺点和问题。我们衷心地希望使用系列教材的教师和学生能够不吝赐教,帮助我们不断地提高系列教材的质量。

曲维枝

2008年12月15日



序 言

日新月异的技术发展及应用变迁不断给信息系统的建设者与管理者带来新的机遇和挑战。例如，以 Web 2.0 为代表的社会性网络应用的发展深层次地改变了人们的社会交往行为以及协作式知识创造的形式，进而被引入企业经营活动中，创造出内部 Wiki (Internal Wiki)、预测市场 (Prediction Market) 等被称为“Enterprise 2.0”的新型应用，为企业知识管理和决策分析提供了更为丰富而强大的手段；以“云计算”(Cloud Computing) 为代表的软件和平台服务技术，将 IT 外包潮流推向了一个新的阶段，像电力资源一样便捷易用的 IT 基础设施和计算能力已成为可能；以数据挖掘为代表的商务智能技术，使得信息资源的开发与利用在战略决策、运作管理、精准营销、个性化服务等各个领域发挥出难以想象的巨大威力。对于不断推陈出新的信息技术与信息系统应用的把握和驾驭能力，已成为现代企业及其他社会组织生存发展的关键要素。

根据 2008 年中国互联网络信息中心 (CNNIC) 发布的《第 23 次中国互联网络发展状况统计报告》显示，我国的互联网用户数量已超过 2.98 亿人，互联网普及率达到 22.6%，网民规模全球第一。与 2000 年相比，我国互联网用户的数量增长了 12 倍。换句话说，在过去的 8 年间，有 2.7 亿中国人开始使用互联网。可以说，这样的增长速度是世界上任何其他国家所无法比拟的，并且可以预期，在今后的数年中，这种令人瞩目的增长速度仍将持续，甚至进一步加快。伴随着改革开放的不断深入，互联网的快速渗透推动着中国经济、社会环境大步迈向信息时代。从而，我国“信息化”进程的重心，也从企业生产活动的自动化，转向了全球化、个性化、虚拟化、智能化、社会化环境下的业务创新与管理提升。

长期以来，信息化建设一直是我国国家战略的重要组成部分，也是国家创新体系的重要平台。近年来，国家在中长期发展规划以及一系列与发展战略相关的文件中充分强调了信息化、网络文化和电子商务的重要性，指出信息化是当今世界发展的大趋势，是推动经济社会发展和变革的重要力量。《2006—2020 年国家信息化发展战略》提出要能“适应转变经济增长方式、全面建设小康社会的需要，更新发展理念，破解发展难题，创新发展模式”，这充分体现出信息化在我国经济、社会转型过程中的深远影响，同时也是对新时期信息化建设和人才培养的新要求。

在这样的形势下，信息管理与信息系统领域的专业人才，只有依靠开阔的视野和前瞻性的思维，才有可能在这迅猛的发展历程中紧跟时代的脚步，并抓住机遇做出开拓性

的贡献。另一方面,信息时代的经营、管理人才以及知识经济环境下各行各业的专业人才,也需要拥有对信息技术发展及其影响力的全面认识和充分的领悟,才能在各自的领域之中把握先机。

因此,信息管理与信息系统的专业教育也面临着持续更新、不断完善的迫切要求。我国信息系统相关专业的教育已经历了较长时间的发展,形成了较为完善的体系,其成效也已初步显现,为我国信息化建设培养了一大批骨干人才。但我们仍然应该清醒地意识到,作为一个快速更迭、动态演进的学科,信息管理与信息系统专业教育必须以综合的视角和发展的眼光不断对自身进行调整和丰富。本系列教材的编撰,就是希望能够通过更为系统化的逻辑体系和更具前瞻性的内容组织,帮助信息管理与信息系统相关领域的学生以及实践者更好地掌握现代信息系统建设与应用的基础知识和基本技能,同时了解技术发展的前沿和行业的最新动态,形成对新现象、新机遇、新挑战的敏锐洞察力。

本系列教材旨在体系设计上较全面地覆盖新时期信息管理与信息系统专业教育的各个知识层面,既包括宏观视角上对信息化相关知识的综合介绍,也包括对信息技术及信息系统应用发展前沿的深入剖析,同时也提供了对信息管理与信息系统建设各项核心任务的系统讲解。此外还对一些重要的信息系统应用形式进行重点讨论。本系列教材主题涵盖信息化概论、信息与知识管理、信息资源开发与管理、管理信息系统、商务智能原理与方法、决策支持系统、信息系统分析与设计、信息组织与检索、电子政务、电子商务、管理系统模拟、信息系统项目管理、信息系统运行与维护、信息系统安全等内容。在编写中注意把握领域知识上的“基础、主流与发展”的关系,体现“管理与技术并重”的领域特征。我们希望,这套系列教材能够成为相关专业学生循序渐进了解和掌握信息管理与信息系统专业知识的系统性学习材料,同时成为知识经济环境下从业人员及管理者的有益参考资料。

作为普通高等教育“十一五”国家级规划教材,本系列教材的编写工作得到了多方面的帮助和支持。在此,我们感谢国家信息化专家咨询委员会及高等学校信息管理与信息系统系列教材编委会专家们对教材体系设计的指导和建议;感谢教材编写者的大量投入以及所在各单位的大力支持;感谢参与本系列教材研讨和编审的各位专家、学者的真知灼见。同时,我们对电子工业出版社在本系列教材编辑和出版过程中所做的各项工作深表谢意。

由于时间和水平有限,本系列教材难免存在不足之处,恳请广大读者批评指正。

高等学校信息管理与信息系统
专业系列教材编委会

2009年1月



前 言

从 20 世纪 90 年代开始,由于全球经济开始从工业经济阶段逐步演进到知识经济阶段,人们对知识在经济发展乃至社会进步中的作用日益重视。知识管理问题也逐步提到了议事日程上来。

知识应用的历史是极其久远的,当人们从以狩猎与采摘为生发展到农耕经济,就已经开始获得和积累生产的知识,但是这些知识是以经验为基础而只能在极为狭窄的范围内传播。进入工业经济阶段之后,科学技术知识渐成体系,对经济发展起到了重要的作用,造纸和印刷术的发展使得知识能够加以保存和传播,也使得知识的作用和影响日益增大。信息技术的迅速发展使科学技术与管理知识的作用更加显著。知识作为一种重要的资源和生产要素,催生了一批高技术产业,推动了传统产业的现代化。由于知识本身的生产与传播独立于物质生产之外,因而成为一种新的专门事业。人们认为已经开始进入知识经济阶段。

知识的创新和广泛应用,引起了生产方式、生活方式、交往方式乃至思维方式的重大改革,人们需要更加自觉地认识和发挥知识的作用。知识作为一种无形资产,它的应用和创新需要自觉地加以管理。由于知识的高度抽象性、流动性和创新性,它的管理需要专门加以研究。因此,一个新的管理领域——知识管理就应运而生。

在总结实践经验和进行跨学科研究的基础上,知识管理逐步形成了独立的学科。在各高等学校管理类的本科与研究生的教学计划中也开始设置知识管理的课程。一般是从开设选修课起步的。

高等学校中的信息管理与信息系统专业培养的是适应社会信息化发展需要的通用信息管理人才,在他们的职业生涯中,会越来越多地涉及知识管理工作。为了使未来的信息管理人才掌握知识管理的基础知识,近年来知识管理课程已经开始列入国内各院校的信息管理与信息系统专业的教学计划,并且已经有多种教材问世。这些教材虽然出发点和视角不同,繁简各异,但都比较系统地阐述了知识管理的基本原理和方法。它们的出版为知识管理的理念、方法和技术的普及起到了推动作用。

在普通高等教育“十一五”国家级教材规划中,“信息化与信息社会”系列丛书之高等学校信息管理与信息系统专业系列教材是其中的组成部分。这一系列教材旨在体系设计上较全面地覆盖新时期信息管理与信息系统专业教育的各个知识层面。知识管理是信

息系统重要的应用领域之一，因此在这一教材系列中列入了这本题为《信息与知识管理》的选题。

在知识管理的教材中，加入一定的信息管理内容是一种新的尝试。我们认为：由于信息是知识（特别是可以言传的显性知识）的载体，因此知识的获取、组织、处理、传递以及知识的应用和创新，都离不开信息方法和工具的应用。在学习知识管理课程之前，学生应该具备一定的信息管理知识。由于信息管理与信息系统专业是由原先许多专业合并而成的，而且原来各专业的侧重面不同，有的侧重系统的开发和运用（原来的管理信息系统专业），有的侧重于信息资源管理（原来的情报专业等信息资源管理类专业），因而各类专业的教学计划也有所不同。根据我们的理解，现在这套教材主要是针对前一类专业的，因此学习知识管理更着重于知识管理系统的开发与运用，也就是更侧重于知识管理中的信息技术应用问题。由于这类专业有关信息管理（特别是信息资源管理）的内容散见于其他各类课程，为了给知识管理提供一些概括性的信息管理知识，本书首先简要地介绍了信息管理的一些基本概念，然后将信息管理与知识管理具有共性的一些概念与方法尽可能合在一起进行讲述。但全书的主要内容和针对性还是知识管理。

本书共有 12 章，各章的内容如下：

第 1 章“信息与知识”介绍了信息与知识的基本概念，它们的分类与特征，以及它们的作用和影响。探讨了知识与创造力经济的特征和影响。

第 2 章“信息管理概述”是对信息管理的概括介绍，阐述了信息管理的含义，从信息资源管理、信息系统管理、信息组织与人员的管理、信息服务诸方面，介绍了信息管理的内容，强调了信息管理中人的因素的重要性。

第 3 章“知识管理概述”是对知识管理的概括介绍。该章以知识系统工程的观点，从知识资源、知识过程、知识组织与人员、知识系统的基础设施、知识经营与知识服务、知识的社会文化环境等各方面，阐明知识管理的特点。

在概括介绍了信息管理与知识管理之后，后面的几章基于信息与知识的共性和相关性，将两者结合起来，介绍其处理与管理方法和工具。

第 4 章“信息与知识资源及其采集”从信息资源与知识资源的对比出发，着重介绍知识资源的含义及特征，以及知识资源的分类。介绍了知识资源采集途径，以及知识搜索等技术。

第 5 章“知识表示”介绍了知识的表示方法，着重介绍了不确定性知识与非结构化信息与知识的表示方法。

第 6 章“信息与知识的组织”介绍了信息组织过程中的几类方法，然后，重点阐述了知识组织的方法，包括语义词典、概念图、主题图、本体和语义网等。

第 7 章“信息与知识的检索”介绍了信息检索的概念及经典的检索模型，知识检索

的概念、机理和评价标准,重点介绍了知识检索模型及基于知识的检索模型和检索结果的可视化方法与主要技术。

第8章“数据挖掘和知识发现”介绍了数据挖掘和知识发现这门新兴的信息与知识处理技术,包括数据挖掘、文本挖掘和 Web 挖掘方法。

第9章“知识系统的技术体系结构”介绍了知识系统中的技术工具和方法的组成与相互间的关系,包括知识系统中的人-机分工与协作,以及信息网络、博客与维基、信息门户与知识门户、知识泵与个性化推荐系统、群件与计算机支持协同工作、Agent 技术等工具和方法。

后面3章则针对知识系统的工作过程、知识系统的开发与典型应用进行了阐述。

第10章“知识系统的工作过程分析”介绍了知识系统在运作层次上的工作过程,包括知识需求与审计、知识采集与编码、知识吸收与利用、知识保存与保护、知识传播与共享、知识转化与生成、知识学习与重用、知识集成与创新等过程。

第11章“知识管理的典型应用”介绍了个人知识管理、项目管理中的知识管理、企业知识管理、政府知识管理的含义与实施。

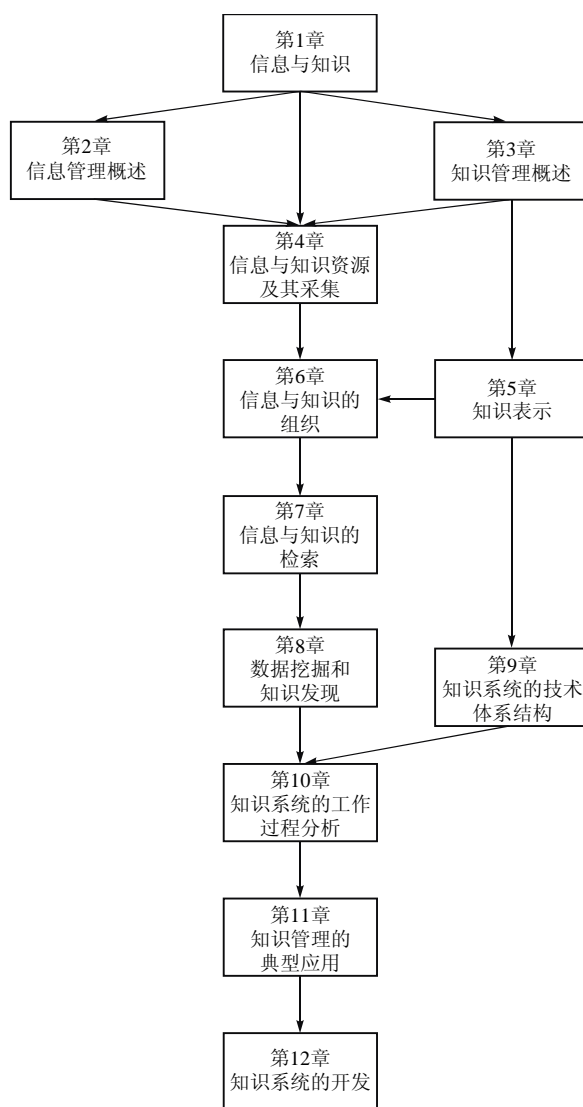
第12章“知识系统的开发”讲述的是知识系统开发的战略考虑和实施步骤。该章对信息基础设施的组成进行了介绍。还介绍了知识流程外包的特点和内容。最后介绍了一类隐性的知识管理系统——学习型组织。

全书的结构与各章之间的关系如下页图所示。

我们认为,学习知识管理课程,首先要掌握有关知识管理的一些基本理念,知道知识管理是“做什么”的,以及知识管理的作用和影响。然后掌握“怎么做”的一些方法和思路。由于本书主要针对以信息系统的开发与运用为主的专业,因此对知识管理系统以及系统中的主要工具做了较多的阐述,但是知识管理的另外一个侧面,即有关知识管理的组织与个人行为问题也是很重要的。本书在第3章中做了初步探讨,希望读者不要因为篇幅较少而不予重视。

对于在学习知识管理课程之前已经学过信息管理概论课程的,在使用本教材时可以略去第1章的前一半和整个第2章。对于学时较少而只希望了解知识管理概貌的课程(例如,其他专业的选修课和通识性课程),可以只用第1、2、3、10、11、12这几章。教材内容可以加以灵活组织和裁减。对于一些培训班来说,学习内容也可以这样选择,甚至可以只使用第1、2、3章作为讲座内容;而对于一些进行创新和创业的人士,选读这几章对他们的事业也会有所帮助。

本书第1、2、3、12章由王众托编写,第4、6、7、10章由吴江宁编写,第5、8、9、11章由郭崇慧编写。

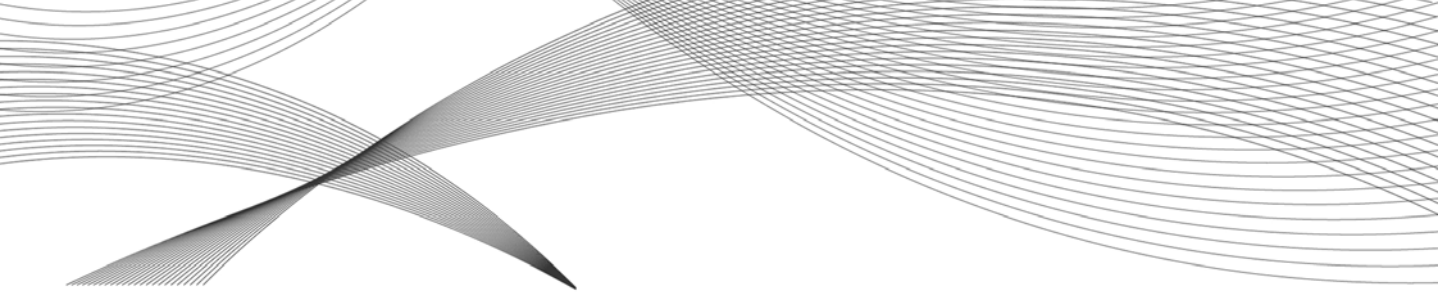


本教材的编写是在编委会统一安排和指导下的进行的。我们衷心感谢编委会各位领导、大连理工大学、管理学院以及系统工程研究所的领导和知识管理课题组对于本书编写提供的工作条件。衷心感谢国家自然科学基金委员会长期的支持和提供的重点课题（79630010 与 70431001）和重大国际合作课题（70620140115）的资助，使我们能够较早地进入这一领域。本教材从选题、建构、写作到完稿一直得到许多专家学者的关注和支持，从他们的著作和谈话中也得到了很多教益。我们特别要感谢我国信息管理学科的创建人之一、著名的管理信息系统专家黄梯云教授，作为本书的主审者，他不但对全书的组织架构给以指导，而且细心地审读了全书文稿，指出了内容和文字中的疏漏和有待改进之处，他的认真负责精神使我们深为铭感。我们还要衷心感谢电子工业出版社的刘宪兰编辑在本书筹划、组织与出版以及作为责任编辑的过程中付出的辛勤劳动。

知识管理是一门新的学科，我们从事这一学科的教学和研究时间还不长，自己的理论水平、实践经验和学术眼界也都还有限，因此书中不足之处在所难免，恳请使用本教材的教师、学员和广大的读者多加批评指正！

编著者谨识

2010 年 3 月



目 录

第 1 章 信息与知识	1
1.1 信息	2
1.1.1 信息的含义	2
1.1.2 信息的特征	3
1.2 信息的性质与类型	4
1.2.1 信息的性质	4
1.2.2 信息的类型	5
1.3 信息的作用和信息化	7
1.3.1 信息的功能与作用	7
1.3.2 信息化及其影响	8
1.4 知识的含义与特征	11
1.4.1 知识的定义	11
1.4.2 知识的特征	12
1.5 知识的类型	13
1.6 知识的作用与影响	16
1.6.1 知识的作用	16
1.6.2 知识经济及其特点	18
1.6.3 创造力经济及其影响	20
思考与讨论题	23
参考文献	23
第 2 章 信息管理概述	25
2.1 信息管理的含义	26
2.1.1 信息管理的概念	26
2.1.2 信息管理的层次	27
2.1.3 信息管理的职能	29
2.2 信息管理过程	30
2.2.1 战略层次的信息管理过程	30
2.2.2 战术层次的信息管理过程	31
2.2.3 运作层次的信息管理过程	32
2.3 信息资源管理	33

2.3.1	对信息资源的进一步认识	33
2.3.2	信息内容管理	34
2.4	信息系统管理	35
2.4.1	信息系统的概念	35
2.4.2	信息系统的体系结构与技术平台	36
2.4.3	信息系统开发管理	38
2.4.4	信息系统运行管理与安全管理	40
2.5	信息组织与人员的管理	41
2.5.1	信息管理的组织结构及其管理	41
2.5.2	信息人员管理	42
2.6	信息服务	43
2.6.1	信息服务的含义	43
2.6.2	信息服务的方式	44
2.7	信息管理中的人的因素问题	45
2.7.1	信息生态	45
2.7.2	信息用户利用信息的心理机制	48
2.7.3	信息素养	50
	思考与讨论题	51
	参考文献	51
第 3 章	知识管理概述	53
3.1	知识管理的概念	54
3.1.1	知识管理的含义	54
3.1.2	知识管理的特点	55
3.2	知识资源、知识资产和知识资本	56
3.2.1	知识资源	56
3.2.2	知识资产	58
3.2.3	知识资本	59
3.3	知识管理的任务与原则	62
3.3.1	知识管理中的主体与客体	62
3.3.2	知识管理的任务	63
3.3.3	知识管理的原则	64
3.4	知识系统与知识系统工程	65
3.4.1	知识系统	65
3.4.2	知识系统工程	67
3.5	知识组织和知识工作者	70
3.5.1	知识组织	70
3.5.2	知识工作和知识工作者	72

3.5.3 知识工作者的特点和对知识工作者的管理	75
3.6 知识运作与管理的技术设施	77
3.6.1 知识运作与管理的技术工具	77
3.6.2 知识库	79
3.7 知识经营和知识服务	80
3.7.1 知识经营	80
3.7.2 知识服务	81
3.8 知识的社会与文化环境	83
思考与讨论题	85
参考文献	85
第4章 信息与知识资源及其采集	87
4.1 信息资源	88
4.1.1 信息资源的概念	88
4.1.2 信息资源的特征	88
4.1.3 信息资源的分类	90
4.2 知识资源	90
4.2.1 知识资源的概念	90
4.2.2 知识资源的特征	91
4.2.3 知识资源模型	92
4.3 知识资源的分类	96
4.4 知识资源的发现与采集	98
4.4.1 知识资源的发现	98
4.4.2 知识资源的采集方式	99
4.5 知识采集技术	100
4.5.1 知识搜索引擎	100
4.5.2 问答系统	103
4.5.3 知识地图技术	104
4.5.4 Agent 技术	105
思考与讨论题	106
参考文献	106
第5章 知识表示	107
5.1 知识表示简介	108
5.2 确定性与不确定性知识	108
5.2.1 确定性知识	108
5.2.2 不确定性知识	109
5.3 确定性知识表示方法	111

5.3.1	逻辑表示法	111
5.3.2	关系表示法	113
5.3.3	产生式规则表示法	113
5.3.4	框架表示法	115
5.3.5	语义网络表示法	116
5.3.6	面向对象表示法	117
5.3.7	本体表示法	118
5.4	不确定性知识表示方法	119
5.4.1	基于概率论的知识表示法	119
5.4.2	基于模糊集理论的知识表示法	120
5.4.3	基于粗糙集理论的知识表示法	122
5.5	非结构化信息的知识表示方法	124
5.5.1	半结构化数据与非结构化数据	124
5.5.2	半结构化数据的知识表示	125
5.5.3	多媒体信息的知识表示	126
	思考与讨论题	127
	参考文献	127
第 6 章	信息与知识的组织	129
6.1	信息组织	130
6.1.1	信息组织的概念	130
6.1.2	信息组织的类型	130
6.1.3	信息组织的原则	130
6.1.4	信息组织的方法	131
6.2	知识组织概述	135
6.2.1	知识组织的概念	135
6.2.2	知识组织的原则	136
6.2.3	知识组织的方式	138
6.3	知识组织的方法	138
6.3.1	语义词典	139
6.3.2	概念图	140
6.3.3	主题图	143
6.3.4	本体	147
6.3.5	语义网	150
6.3.6	特定领域分类法	151
6.3.7	网络分类法	152
6.3.8	分众分类法	154
6.4	知识组织的技术	156

6.4.1 知识加工	156
6.4.2 知识表示	156
6.4.3 知识聚类	156
6.4.4 知识重组	156
6.4.5 知识存储	157
思考与讨论题	157
参考文献	157
第 7 章 信息与知识的检索	159
7.1 信息检索	160
7.1.1 信息检索的概念	160
7.1.2 信息检索的过程	160
7.1.3 信息检索模型	161
7.2 知识检索	173
7.2.1 知识检索的概念	173
7.2.2 知识检索的特征	174
7.2.3 知识检索机理	174
7.2.4 检索效果评价	175
7.3 知识检索模型	176
7.3.1 基于概念图的检索模型	176
7.3.2 基于主题图的检索模型	179
7.3.3 基于本体的检索模型	180
7.3.4 分类检索模型	181
7.3.5 多维认知检索模型	182
7.3.6 分布式检索模型	182
7.4 检索结果可视化	183
7.4.1 呈现方法	183
7.4.2 实现技术	187
思考与讨论题	188
参考文献	188
第 8 章 数据挖掘和知识发现	189
8.1 数据挖掘和知识发现简介	190
8.2 知识发现的基本步骤	191
8.3 数据挖掘的主要任务和基本方法	192
8.3.1 降维与特征提取	192
8.3.2 关联规则	193
8.3.3 分类与回归	195

8.3.4 聚类 and 异常点检测	198
8.4 文本挖掘	200
8.4.1 文本挖掘简介	200
8.4.2 文本数据预处理	201
8.4.3 文本分类	202
8.4.4 文本聚类	203
8.4.5 文本关联分析	203
8.5 Web 挖掘	204
8.5.1 Web 挖掘简介	204
8.5.2 Web 内容挖掘	205
8.5.3 Web 结构挖掘	206
8.5.4 Web 使用挖掘	206
思考与讨论题	207
参考文献	207
第 9 章 知识系统的技术体系结构	209
9.1 知识系统中的人-机分工与协作	210
9.2 知识系统中的技术工具体系	210
9.2.1 信息基础设施与计算机系统	211
9.2.2 信息网络的类型	212
9.2.3 互联网与万维网	213
9.2.4 内联网与外联网	215
9.3 博客与维基在知识管理中的作用	217
9.3.1 博客与知识管理	217
9.3.2 维基与知识管理	218
9.4 信息门户与知识门户	220
9.4.1 信息门户	220
9.4.2 知识门户	221
9.5 知识泵与个性化推荐系统	222
9.5.1 知识泵	223
9.5.2 个性化推荐系统	224
9.6 群件与计算机支持协同工作	226
9.6.1 群件	226
9.6.2 计算机支持协同工作	230
9.7 Agent 及其在知识管理系统中的应用	233
9.7.1 Agent 技术简介	233
9.7.2 基于多 Agent 的知识管理系统	234
思考与讨论题	237

参考文献	238
第 10 章 知识系统的工作过程分析	239
10.1 知识系统工作过程描述	240
10.2 知识需求与审计	240
10.2.1 知识需求	240
10.2.2 知识审计	243
10.3 知识采集与编码	244
10.3.1 知识采集	244
10.3.2 知识编码	246
10.4 知识吸收与利用	247
10.5 知识保存与保护	249
10.5.1 知识存储	249
10.5.2 知识产权保护	251
10.6 知识传播与共享	251
10.6.1 知识传播	251
10.6.2 知识共享	252
10.7 知识转化与生成	254
10.7.1 知识转化模型	254
10.7.2 知识生成环境	259
10.8 知识学习与重用	262
10.8.1 组织学习	262
10.8.2 知识重用	263
10.9 知识集成与创新	265
10.9.1 知识集成	265
10.9.2 知识创新	266
思考与讨论题	267
参考文献	267
第 11 章 知识管理的典型应用	269
11.1 个人知识管理	270
11.1.1 个人知识管理的含义	270
11.1.2 个人知识管理的实施	271
11.1.3 个人知识管理的支持工具	273
11.2 项目管理中的知识管理	276
11.2.1 项目管理的特点	276
11.2.2 项目管理中的知识	277
11.2.3 项目各阶段的知识管理	278

11.2.4 专家知识的获取	280
11.2.5 项目管理中的知识集成	282
11.3 企业知识管理	284
11.3.1 企业知识管理的含义	284
11.3.2 企业知识管理的实施	285
11.3.3 企业知识管理系统	287
11.4 政府知识管理	291
11.4.1 政府知识管理的含义	291
11.4.2 政府知识管理的实施	292
思考与讨论题	294
参考文献	295
第 12 章 知识系统的开发	297
12.1 知识项目与知识系统	298
12.2 知识系统开发的战略考虑	299
12.2.1 知识战略与组织总战略	299
12.2.2 知识战略的组成	300
12.2.3 知识战略的实施	302
12.3 知识系统开发的步骤	303
12.4 知识系统的基础技术设施	305
12.5 知识流程的外包	309
12.6 作为隐性系统的学习型组织	311
12.6.1 组织学习的含义与内容	311
12.6.2 学习型组织的特点	312
12.6.3 实践社群	313
思考与讨论题	315
参考文献	315

第 1 章

信息与知识

内容提要

本章讲述的是信息与知识的基本概念。首先介绍有关信息的概念和定义，以及信息的特征和性质；然后介绍信息的功能与作用，信息化的内容和影响；其次介绍知识的概念和定义，知识的特征与类型，知识的作用与影响，以及知识与创造力经济的意义和影响。

本章重点

- 信息的定义与特征
- 知识的定义与类型
- 知识的作用与影响



1.1 信息

我们先来研究一下信息的含义与特征。

1.1.1 信息的含义

信息是人们当前经常听到和使用的一个概念和名词。在日常生活中，上下班和出行需要了解气象信息和交通信息，操作生产设备需要了解机器和原料信息，购买物品需要了解市场信息，求职需要了解就业信息。至于书刊报纸中的文字图表，广播、电视中的声音图像和互联网上的网页，都无时无刻不在传播信息。但是如果问到信息究竟是什么，倒一时难以得到一个统一而明确的答案。

比较一般的理解，信息是消息的意思。中国古代《三国志》书中就有“正数欲来，信息甚大”的记载，唐代就有“梦断美人沉信息，目穿长路倚楼台”的诗句，这是最早在文字记载中出现的信息词语，其含义就是消息。在日常生活中，信息还常常被理解为数据、信号、新闻、情报、文档等的同义语。从科学概念来看，不同学科从不同侧面给信息下了不同的定义。同时由于信息的重要性和地位不断提高，人们对信息的认识也在不断加深，信息的含义也在不断发展，现在还很难得到一个众所公认的定义。

中国著名的信息科学专家钟义信曾经提出^[1,2]，从信息的本质来说（也就是从哲学的本体论层次来说）：信息是事物存在方式或运动状态以及这种方式和状态的自我表述。这是一个比较有普遍意义的定义。它表明只要事物存在并且运动，就有信息存在。这里所谓事物，既包括物质实体，也包括精神世界。所要表述的既有事物内部的结构，也有它和外部的联系。例如，对一台计算机，有关它的信息既有它的内部结构，也有它和外部的链接。对一个组织（如企业）也是既有组织内部结构的信息存在，也有它与外界的联系信息存在。所谓自我表述，就是说不论信息是不是被某种主体所感知，它总是客观存在而向外界输出的。

从信息的可感知角度（也就是从哲学的认识论层次）来看：信息是主体所感知或所表述的关于该事物的运动状态及其变化方式，包括这种状态或方式的形式、含义和效用。这里所说的主体，可以是人、生物或机器（狭义的主体仅指人）。主体所感知是外部世界向主体输入的信息，而主体所表述的事物运动状态及其变化方式，则是主体向外部世界输出的信息。由于引入了主体，而主体具有感觉能力、理解能力和目的性。因此主体只有真正掌握事物在认识论层次的信息，才能做出正确的判断和决策。同时因为人们研究信息总是希望信息为人类服务，所以与主体相联系的认识论层次的信息，将受到特别的关注。

前面曾经说到，信息常被理解为消息、情报、数据、信号、新闻、文档等的同义语，

信息确实与这些名词概念有联系，但还是有一定的区别。

对于消息来说，它是信息的反映形式，或者说消息是带有一定外壳的信息，信息是消息的实质内容。

情报是有目的、有时效的关于人和事的报道或资料，如科技情报，其中包含了科学技术信息。可以说所有的情报都是信息（因此在有些场合科技情报已经改称为科技信息了），但是不能说所有的信息都是情报。文献、档案和新闻等也都是如此。

至于信息与信号的关系，可以说信号是信息的载体，而信息是信号中包含的内容。

通常所说的数据，有些是以数字形式表示的信息。由于数字式电子计算机的广泛应用，许多信息都是以数字形式存储和处理的，所以往往在日常谈话涉及计算机时说到的数据，有一些实际上已经是数字形式表示的信息了。但是从科学意义上来说，数据与信息还是有原则区别的。这一点本书在后面还要谈到。

1.1.2 信息的特征

为了深化对信息的理解，我们来研究一下信息的某些基本特征^[1,2]。

信息来源于物质，但不是物质本身。它从物质的运动中产生，又可以脱离源物质而寄生于其他物质（媒体），相对独立存在。

信息也来源于精神世界，但又不限于精神领域。信息既然是主体的感知或表述，当然是包含像思想、情绪、意志等精神领域中的东西。但是信息还可能产生于物质客体（如气候等自然现象），所以说不仅限于精神领域。

信息与能量息息相关，但又与能量有质的区别。信息与能量都与实物的运动有关，都是事物运动的状态函数。传输和处理信息需要能量的支持，而能量的转换与利用需要信息来引导，所以两者关系密切。但是两者在质的方面的区别反映在能量提供的是动力来源，做工的本领，而信息提供的是运动状态和对物质与能量运动的指引。控制论的创始人维纳曾经说过：“信息就是信息，既不是物质，又不是能量”^[3]。信息和物质、能量是支撑人类社会发展的三种最基本的资源，也是科学技术中的三个最基本的概念。

信息可以提炼成知识，但信息本身不等于知识。信息虽然可以告诉人们事物运动的状态是什么，以及状态会以什么方式改变，但不具有普遍抽象的品质。只有经过积累、抽象提炼，具有更加普遍、深刻与抽象的品格，才能形成知识。关于信息与知识的关系，本书后面还要详细讨论。

信息是具体的，可以被主体（人、生物、机器）所感知、提取和识别，可以被传递、存储、变换、处理、显示、检索和利用。任何人和其他生物的感觉器官以及机器的感受装置，都是为了感知信息的，人的神经系统和机器的信息处理系统，就是用于处理信息的。正是这种处理，使得信息更加丰富，更有实际用途。

信息可以被复制，可以共享。由于信息可以脱离源事物而相对独立地存在，并附着于其他载体，所以可以无限制被复制、传播或分配给许多用户共享。物质和能量不能被共享，被一个主体占有就不能被另一个主体占有；而信息则不然，可被许多主体共同占有共享。

1.2 信息的性质与类型

我们再来研究一下信息的各种性质与分类方法。

1.2.1 信息的性质

从前面列举的信息基本定义和基本特征，可以导出信息的如下一些基本性质^[1]。

(1) 信息的普遍性：信息是普遍存在的。

只要有事物存在，只要事物在不断地运动，就会有它的运动状态和状态变化的方式，也就存在信息。无论是自然界、人类社会还是人的思维领域，绝对真空和绝对静止是没有的，因此信息普遍存在。

(2) 信息的无限性：在整个宇宙时空中，信息是无限的，即使在有限空间，随着时间的推移，信息也是无限的。

由于宇宙时空中的事物及其变化是无限丰富的，所以产生的信息也是无限的。在有限的空间（如地球），在无限的时间长河中，事物的发展变化也是无限的，因而信息也是无限的。但是这里需要说明的是，信息有实在信息与实得信息的区别，实在信息是无限的，而人类实际得到的信息却只是有限的一小部分。

(3) 信息的依附性：信息必须依附在载体上，不可能单独存在。

信息只有依附在信息载体上，才能传输、保存和转换，载体可以是实物，如书刊、光盘等，也可以是计算机网络中的数码，还可以是电磁波。

(4) 信息的传递性：信息可以在时空中从一点传递到另一点。

由于信息可以脱离母体相对独立，因此可以通过一定方法进行传递。在空间中传递称为通信，在时间上的传递称为存储（存储也可看作是过去和现在、现在和未来的通信）。这就大大缩短了空间的距离和时间的距离。

(5) 信息的可变换性：信息是可以变换的，它可以有不同的载体用不同的方法来承载。

只要保持信息所表述的运动状态以及其变化方式不变，也就是信息内容不变，就可以在载体的不同物质或能量形式之间变换。例如，信息既可用光缆传输，又可用电波无线传输。

(6) 信息的有序性：信息可以用来消除系统的不确定性，增加其有序性。

信息的获得可以消除主体对事物状态及其变化方式的不确定性。

(7) 信息的动态性：一切活的信息都随时间而变化。

由于事物本身是在不断发展变化的，信息当然也随着变化。因而信息有一定的时效性，也有自己的生命周期。

(8) 信息的相对性：对于同一事物，不同的观察者得到的信息量可能不同。

由于不同观察者的观察能力、理解能力和原先掌握的信息量不同，所以对同一事物所获得的信息量也各不相同，这也就是说，实际得到的信息量因人而异。

通过对信息的上面这些性质的认识，可以使人们对过去难以捉摸的信息进一步有了一般性的理解。其实如果循着历史的长河来进行考察，人们对于物质的一般概念的掌握是经历了一个过程的，也是在接触和利用了从岩石到金属再到合成材料的各式各样的物质之后，才产生了物质的一般概念。对能量也是这样，开始时觉得难以捉摸，在接触和利用了动能、热能、电能和核能等各式各样的能源之后才形成了对能源的一般概念。对信息来说，不同的人们接触了不同类型的信息，开始也是从最熟悉的信息形式开始认识的，有的把信息看作是消息，有的看作是情报等。后来接触的信息形式多了，也就对信息的各种性质和特征有了比较全面的了解，逐步建立了信息的一般概念。

1.2.2 信息的类型

信息可以按照不同的属性原则，进行不同类型的划分。

1. 按照信息来源的领域划分

按照信息来源的领域划分，可以分为：

(1) 天然信息。这是来自大自然的信息，如天文信息、海浪信息等。人们获取这类信息，为的是适应自然和在一定条件下改造自然。

(2) 人工自然信息。人工自然界是依靠自然环境和资源由人类劳动建造出来的物质世界（例如，水电站、工厂、机床、计算机网络等），它与天然的自然界（也可以称为天工自然界）不同，也不同于人类社会，但还是物质世界。人工自然信息是来自人类改造过的自然界和制造出来的器物的信息，如水库的水位，电气设备上指示灯的亮光。

(3) 社会信息。它是人类社会自身的活动所产生的信息，主要是人类联系、沟通的信息，如报刊上对一些会议活动的报道。

(4) 思维信息。这是涉及个人思维活动的信息，如人的情绪、人的表情、人的创意、人所做出的判断等。现在已经可以通过仪器从人脑得到一部分与此有关的信息。

上述按照来源的领域划分中，后面三种都涉及人的活动，因此可通称为人工信息。它与自然信息相并列，相对应。

2. 按照信息感知程度划分

按照信息感知程度划分, 可以分为:

(1) 实在信息。这是事物本身固有的信息, 它只取决于事物本身的运动状态和状态变化的方式, 与主体无关。它还有一个名称: 自在信息。

(2) 实得信息。这是主体已经获得的信息。它不仅具有事物本身的客观属性, 也取决于主体的某些属性, 如感知能力。

(3) 先验信息。这是某一主体在感受事物之前就已经具备的有关该事物的信息。这种信息不但与主体的感知能力有关, 还与主体的另一些主观因素如个人经验、认知能力有关。

3. 按照主体的认识层次划分

按照主体的认识层次划分, 同一信息, 可能包含下面几个层次的内容:

(1) 语法信息。这是对信息认识的第一层次, 只反映事物的存在方式和运动状态, 不考虑信息的内涵, 具有客观的本性。例如, 汽车鸣笛, 声波这种形式就是语法层次信息。

(2) 语义信息。这是第二个层次, 反映的是事物存在方式和运动状态的含义。例如, 人们听到鸣笛声时, 就知道汽车走近了, 这时接受的既有语法层次的又有语义层次的信息。

(3) 语用信息。这是第三个层次, 反映的是事物存在方式和运动状态的效用。同样的信息, 对不同主体的效用是不一样的。

例如, 两个人走在路中央, 后面汽车在鸣笛。鸣笛的声波传递的是语法信息, 是一种客观的信息(具备一定的信息形式)。两个人都从鸣笛声知道是汽车来了, 这时两人都获得了语法信息以及语义信息。但两个人中一个人事先约定将有车来接他, 这个信息告诉他: 这是接他的汽车到了; 另一个人只是根据一般常识做出的判断, 认为是汽车在示警, 要他让路。同样信息对两个人的效用是不一样的, 因为他们事先已经具备的信息和理解不一样。

4. 按照信息与主体的意识有无关系划分

按照信息与主体的意识有无关系划分, 可以分为:

(1) 客观信息。这是存在于作为主体的人的意识之外的信息。

(2) 主观信息。这是依赖于人的主观意识的信息, 信息中夹杂有出自个人意识活动的部分。

5. 按照信息与时间的关系划分

按照信息与时间的关系划分, 可以分为:

(1) 静态信息。由于事物不可能永远不发生变化, 所以没有绝对静态的信息。一般

所说的静态信息，在相当一段时间内是不随时间变化的，如历史文献和天文记录等。

(2) 动态信息。这类信息是随时间发生变化的，有些信息的时效性是很强的。

6. 按照信息是否可以用数量来描述划分

按照信息是否可以用数量来描述划分，信息可以分为：

(1) 定量信息。这是可以量化的信息。

(2) 定性信息。这是不可以量化，只能由自然语言表述其特性的信息。

除了上面提到的分类方法外，还有一些分类法，如宏观信息与微观信息，直接信息与间接信息，原始信息与二次信息等，这都与日常生活中的划分法一致，与我们理解信息特点无关，就不必再列举了。

1.3 信息的作用和信息化

信息究竟有什么作用，又怎样对社会产生影响，这是本节讨论的内容。

1.3.1 信息的功能与作用

信息的功能是很多的。不同的信息学科领域对信息功能的理解是不一样的，下面列举一些带有普遍性的功能。

1. 信息的显示、表述功能

按照信息的基本定义：“信息是事物存在方式或运动状态以及这种方式和状态的自我表述”，信息的功能首先是显示、表述事物状态和变化过程。

2. 信息的联系功能

通过信息，把不同的对象联系了起来（如不同的计算机通过信息联系起来），或者把主体与认识对象联系起来。其中最重要的是使主体通过信息与环境联系起来，感受环境的变化。作为主体的生物要想生存，必须适应环境，这首先要获取环境信息。生物的进化在于获得信息的能力增强而获得的信息日益增多。

3. 信息的保持（存储）功能

进化到人类，已经不仅能获得各方面的信息，包括自然与社会的信息，由于信息的可存储性，人已经能对信息进行记忆和加工，以积累经验。

4. 信息的可转换和可深加工的功能

信息的可转换和可深加工的功能，使得人类能够在原来掌握信息的基础上产生新的想法，指导新的行动，不仅适应自然和社会，而且在一定程度上可以改造自然和社会。

人类利用信息的上述功能，在自己的生活实践中，使信息发挥出下面这些作用。

1. 信息的认识作用

人类通过获取信息来认识事物，最初仅仅是通过自己的感受器官（如肉眼）直接认识事物，后来创造了一些工具（如望远镜、显微镜）来获取信息，扩大了认识范围和认识的深度。

2. 信息的传播与沟通作用

信息传播是人类早期的活动之一，没有传播就无法与其他人联系，无法进行人际沟通。社会活动是建立在信息传播和沟通基础之上的。

3. 信息的组织作用

任何组织的形成必须以信息的沟通和共识的形成为先决条件，这时信息起到了重要作用。组织的生存和发展需要组织成员的协同一致，而信息起到了导向和指引作用。一个系统的组成可以是外界组织的，也可以是系统自组织的，两者都需要信息发挥作用。

4. 信息的控制与管理作用

控制（无论是人工控制还是自动控制）和管理具有共性，都是有目的的干预。要想实施控制或管理，必须了解对象的情况，这要依靠信息的采集和加工。当需要采取行动时，需要通过信息来协调各部分的动作。

随着科学技术的进步，人类创造发明了许多信息工具，创建和发展了信息技术，在以下四个方面扩展了人的信息处理能力。

（1）在获取信息方面：单靠人的信息感觉（包括视觉、听觉、嗅觉、味觉和触觉）器官来获取信息是有局限性的，后来发明了许多感测技术工具，扩展了人的感受范围和深度。

（2）在传递信息方面：单靠人与人之间的口头联系范围也太狭窄了，一些通信工具的发展和广泛利用，缩短了空间和时间的距离，原来需要长久才能到达的信息变得瞬间就可以传递到。信息使得人们感到世界变小了，感到时间变快了。

（3）在加工信息方面：在记忆、联想、分析、综合、判断和决策等方面，单靠人脑的信息处理能力无论是速度和深度也都是太低了，像计算机这类工具的发明大大扩展了人的某些智力活动能力。

（4）在施用信息方面：人的效应器官（语言用口，行走用脚，操作用手）都有一定限度，而基于信息的控制技术可以扩展这些能力。

所有这些信息工具的发明使得人在生产和生活方面得到很大的支持，推动了经济的发展，使社会的信息化得以实现。

1.3.2 信息化及其影响

所谓信息化，就是在经济、科技和社会各个领域，广泛应用现代信息技术，有效开

发利用信息资源,建设先进的信息基础设施,发展信息技术和产业,不断提高综合实力和竞争力,加速现代化进程,使信息产业在国民经济中的比重逐步上升的过程。

信息化的物质技术基础是信息技术。信息技术是能扩展人的信息功能的技术,主要是指利用电子计算机和现代通信手段实现获取信息、传递信息、存储信息、处理信息、显示信息和分配信息等的相关技术,也就是充分开发和利用信息资源的技术。在工业发达国家,一般都把信息当作社会生产力发展和国民经济发展的重要资源。

信息技术主要包括以下几方面技术:

(1) 感测与识别技术。它的作用是扩展人获取信息的感觉器官功能。它包括信息识别、信息提取、信息检测等技术。信息识别包括文字识别、语音识别和图形识别等。

(2) 信息传递技术。它的主要功能是实现信息快速、可靠、安全的转移。各种通信技术都属于这个范畴。广播技术也是一种传递信息的技术。

(3) 信息处理与再生技术。信息处理包括对信息的编码、压缩、加密、分析和变换等。在对信息进行处理的基础上,还可形成一些新的更深层次的决策信息,生成知识。信息的处理是依靠现代电子计算机的卓越功能完成的。

(4) 信息应用技术。这是信息过程的最后环节。它包括控制技术和智能支持技术等。其中现代计算机技术和通信技术是信息技术的两大支柱。

为了推进信息化,需要发展信息产业。信息产业一般指以信息为资源,信息技术为基础,进行信息资源的研究、开发和应用,以及对信息进行收集、生产、处理、传递、存储和经营活动,为经济发展及社会进步提供有效的、综合性的生产和经营活动的行业。信息产业包括信息装备制造业和信息服务业。

信息化是社会经济的发展从以物质和能量为经济结构的重心,向以信息和知识为经济结构的重心转变的过程。在这个过程中,不断地采用现代技术装备国民经济各部门和社会各领域,从而极大地提高了社会劳动生产力。

当前无论是发达国家还是发展中国家,都把加快推进信息化作为新世纪的战略任务,并采取了許多重要措施:

(1) 着眼于总体和长期可持续发展,制定统一的国家信息化发展规划,引导和指导全社会的力量参与信息化建设。

(2) 组建高层次的信息化协调和管理机构,强化政府对信息化建设的宏观组织协调功能。

(3) 调整原有的政策法规,制定新的法律法规,为信息化建设创造良好的政策环境。

(4) 及时发布信息化的项目指南,对企业投资进行信息指导,并加大基础性、关键性和引导性信息化项目的投资力度。

近年来中国的信息化工作已经全面展开,并取得很大进展。无论是信息装备制造业

还是信息服务业都已形成规模。在中国社会经济发展的进程中,既面临着良好的发展机遇和有利条件,又面临着紧迫的挑战和不利因素。如果能够牢牢地把握信息化这个千载难逢的机遇,发挥后发优势,就可以实现经济跨越式发展以及社会各方面工作的进步。

要建立的国家信息化体系包括下面六个要素:

(1) 信息资源。信息资源是国民经济和社会发展的战略资源,信息资源开发和利用的程度是衡量国家信息化水平的一个重要标志。

(2) 国家信息网络。国家信息网络是信息资源开发利用和信息技术应用的基础,是信息传输、交换和资源共享的重要手段。特别是互联网,遍布各地,现在要逐步做到网络无处不在。

(3) 信息技术应用。信息技术应用是指要把信息技术广泛应用于经济和社会各个领域。这直接关系到国民经济整体素质、效益和人民生活质量的提高,是国家信息化建设的重要任务。

(4) 信息科学技术和产业。这是中国进行信息化建设的基础。信息化建设要立足于具有自主知识产权的技术和国产装备,这不仅是国家经济发展的需要,也是国家安全的需要。

(5) 信息化人才。我们需要建立一支结构合理、高素质的研究、开发、生产和应用的队伍,以适应国家信息化建设的需要。人才队伍对其他各个要素的发展速度和质量,有着决定性的影响,是信息化建设的关键。

(6) 信息化政策、法规和标准。需要建立一个促进信息化建设的政策、法规环境和标准体系,规范和协调各要素之间的关系,以保障国家信息化的快速、有序和健康发展。

信息技术和信息产业正在成为许多国家经济增长的主要推动力量,不少发达国家信息的增长率均高于国民生产总值增长率,成为新的经济增长点。信息技术和信息产业同时又在改变传统的生产和经营方式甚至生活方式。传统经济的模式正在被新经济的模式取代。发达国家产业结构的升级和经济结构的转型在更大的领域中展开,整个经济正在进入知识经济阶段。

信息化的高速发展大大推动了经济和社会发展^[4]。信息化对世界经济推动的作用令人瞩目。电子商务的问世,直接映射出信息化对世界经济的贡献。20世纪70年代电子商务初露端倪,表现形式仅为企业之间的电子数据交换。20世纪90年代,逐步发展为实时交易。21世纪初,已经达到面向对象综合应用这一较高水准。迅猛发展的全球手机与移动通信产业,在新的信息应用技术催化下,派生出有别传统经济的另类信息经济市场。在互联网基础上正在建立所谓物联网,不但使人与人之间通过网络进行联系,而且人与物之间以及物与物之间也建立了联系,对人类的生产和服务活动提供了极大的方便。

信息化给世界政治带来的冲击具有全球性。信息的网络化状态,突破了传统意义的

疆界限制。任何信息只要能被送上互联网加以传播,便可引发全球的关注。信息时代是网络民主的时代,通过互联网,网民可以帮助政府了解真正的社情民意,提高政府的服务水平。

信息化的制胜效用在战争实践中得到验证,已经深深触动了世界各国。不少国家以此为契机,加快了军事信息化的步伐。

社会信息化的推进,对传播和发展人类科学文化所起的变革作用或影响极其巨大。这一点在学习和教育领域表现得格外突出。电子化的学习手段和网络化的远程教育,给人们提供了广泛汲取知识的自学习方式和不同于传统模式的受教育途径。网络化与虚拟化、公开化与国际化、自主化与多样化,已经成为当代历史进程中全球教育发展的三大趋势。

对于个人而言,信息技术及其应用已经使人们不仅感受到“手的延长”,甚至得到“脑的扩展”。信息技术多样化的应用在各个领域得到普及。信息设备的硬件成本不断降低,使之广泛进入家庭,正成为人们顺手可得的身边工具。在它的支持下,人类个体的创造潜能获得从未有过的激发和最大程度的释放。普通人单凭一台计算机也能很好地发挥出自身创造性,并通过打印机、移动式存储设备或随处可及的网络系统,把个人创造成果广为传播,以实现自然人个体对整个人类社会物质及精神财富积累过程的参与。

1.4 知识的含义与特征

现在转到对知识的研究,首先探讨知识的含义与特征。

1.4.1 知识的定义

知识像信息一样,也是人们日常生活中经常使用的词语和概念,人们进行学习就是为了获得知识;在工作中为了解决问题、进行决策,需要使用知识;创新则是生成新知识。但是要想对知识做出一个众所公认的确切定义还是很困难的。由于知识是一个内涵丰富、外延广泛的概念,对于知识的定义,不同学科有不同的说法,同时知识又是一个发展中的概念,不同历史时期人们也会有不同的理解。下面不妨列举一些常见的知识定义。

中国的《汉语大词典》对知识的解释是“人类认识自然和社会的成果或结晶”。

《现代汉语词典》把知识定义为“人们在改造世界的实践中所获得的认识和经验的总和。”

马克思主义哲学认为,知识的本质在于它从社会实践中来,社会实践是一切知识的基础和检验知识的标准。

从经济学角度来看,知识是人类劳动的产品,是具有价值与使用价值的人类劳动产

品。知识是一种生产要素，是一种无形资产。

从信息的角度来看，知识是同类信息的累积，是为有助于实现某种特定目的而抽象化和一般化了的信息。但也有人认为，知识乃是人对一系列相关信息所产生的反应，知识存在于人而不存在于信息集合之中。

在探讨知识管理时，我们认为，可以应用中国国家科技领导小组办公室在《关于知识经济与国家基础设施的研究报告》中，对知识经济中的知识做出过的定义：

“知识乃是经过人的思维整理过的信息、数据、形象、意象、价值标准以及社会的其他符号化产物，不仅包括科学技术知识（这是知识中的重要组成部分），还包括人文、社会科学的知识，商业活动、日常生活和工作中的经验和知识，人们获取、运用和创造知识的知识，以及面临问题做出判断和提出解决方法的知识。”

这个定义有利于日后有关知识管理的讨论，但是它涉及的面较广，必须与后面要阐述的知识性质与特点以及分类相结合来理解。

1.4.2 知识的特征

前面列举的知识各种定义，从不同侧面说明了知识的性质和特点。我们不必刻意去追求一种统一的定义，不妨从下面几方面去理解知识的本质。

(1) 知识是人类在实践中获得的有关自然、社会、思维现象与本质的认识的总结。

(2) 知识是具有客观性的意识现象，是人类最重要的意识成果。

(3) 一般来说，信息是知识的载体。其中的一部分需要借助于物质载体才能保存与流通。

(4) 从静态来说，知识表现为有一定结构的知识对象；从动态来说，知识是在不断地流动中产生、传递和使用的。

最后这一点表明，知识既可看作一种对象，又可看作是一种过程，就像人们对于光的认识，既可以从它的微粒性着眼，又可以从它的波动性着眼一样。也可以这样说：在任何时刻，都有知识存量；在任何时间段内，都有知识流量。

知识作为人类的一种特定的精神产品，具有下列特征：

(1) 知识是可以存储的，有一些存储于信息的载体（如书刊、半导体或磁性存储器）之中，有一些存储在人的头脑之中，载体是多式多样的。

(2) 知识是可以越过时空传递的，过去的知识可以流传到现在，一地的知识可以传递到其他地方。

(3) 知识是可以传播和分享的，一个人掌握了某种知识，不排除其他人也可同时掌握这些知识，而物质产品就不具备这一特征。

(4) 知识是可以再生的，具有无限复制扩散的能力。

(5) 知识是可以重复使用的, 不存在损耗。

(6) 知识具有相对性, 某一项知识只在特定的时期内或在特定的环境下, 对特定的对象才是有效的。例如, 某一年春节铁路运输的车次安排和售票临时规定之类的知识, 只在当年才有效, 过期就失去了它的有效性。因此对某一项知识来说, 也有它的生命周期。

这里还有必要对数据、信息与知识之间的区别和相互之间的关系加以讨论。数据是经过组织的数字、词语、声音、图像, 是事物属性及其相互关系等的抽象表示。由于经过抽象, 所以数据本身并不具备意义。信息则是以有意义的形式加以排列和处理的数据, 是有目的、有意义、有用途的数据。而知识是对信息进行深加工, 经过逻辑或非逻辑思维, 认识事物的本质而形成的经验与理论。

1.5 知识的类型

知识按照不同的分类法, 可以分成不同的类型。这里只列举一些在后面的讲述中直接有用的几类划分方法。

按照经济合作与发展组织(OECD)报告的分类法^[5], 为了有利于经济分析, 把知识分成如下四类:

(1) “知道是什么”的知识(Know-what), 指的是关于历史事实、经验总结、统计数据的知识, 这类知识与通常所说的信息很难区分。

(2) “知道为什么”的知识(Know-why), 指的是那些自然、社会和人的思维运动的法则和规律的科学知识。这些知识与经济活动的联系并不是直接的, 但在多数产业中, 需要它支持技术的发展和产品与工艺的进步。

(3) “知道怎么做”的知识(Know-how), 是关于技能与诀窍方面的知识, 指的是怎样做某件事情的能力。它与生产相关, 但也与经济领域中像经营管理等其他很多活动相关。熟练工人操作复杂的机器也需要这类知识。应该提到的是, 不仅从事实工作的人需要这类知识, 科学家也需要这类知识。

(4) “知道是谁”的知识(Know-who), 包括关于谁知道什么以及谁知道怎样做什么的信息。特别是它还包括与有关专家形成的特殊社会关系, 以便有可能获得并有效利用这些专家的知识。

前两种知识是可以用语言文字表达的, 在书籍、杂志、报纸、设计文件、图纸等载体中包含的就是这一类知识。后两种知识中相当一部分是无法或很难用语言文字表达的。

正是由于知识在能否用语言文字表达上的区别, 可以把知识分成^[6, 7]如下两类:

(1) “显性知识”或称“言传性知识”(Explicit Knowledge);

(2) “隐性知识”或称“意会性知识”(Tacit Knowledge)。

前一类知识是可以语言文字表达的,在书籍、杂志、报纸、设计文件、图纸等载体中包含的就是这一类知识。由于科学技术的发展,记录和表达的方式也越来越多,如文字、语言、数据、图形、图像、视频等,除了书刊、图纸外,磁性、半导体和光电存储设备等都是新的物质载体。对一个企业或其他组织来说,上述知识都是宝贵的知识资源。这些资源由于可以用语言文字传递(所以称为言传型知识)、交流和保存,其作用和影响是比较明显的。由于这类知识可以编码输入计算机,所以也有人称为可编码的知识(Codified Knowledge)。

后一类知识由于包含了经验、技巧和诀窍,是要靠实践摸索和体验来获得的,可意会而不可言传(因此称为意会性知识更能达意一些)。国外喜欢引用哲学家波兰尼对这类知识的研究成果^[8],他曾提出过一个基本原理:“我们所知道的总是比我们说出来的多”。其实早在两千多年前,中国解释《易经》的《易传》中就说过:“书不尽言,言不尽意”,也就是说,我们不可能把说得出来的都写出来,不可能把意会的东西都说出来。而这类说不出的意会性的经验、体会就是隐性(意会性)的知识^[6]。它们之间的关系可用图 1.1 来表示。

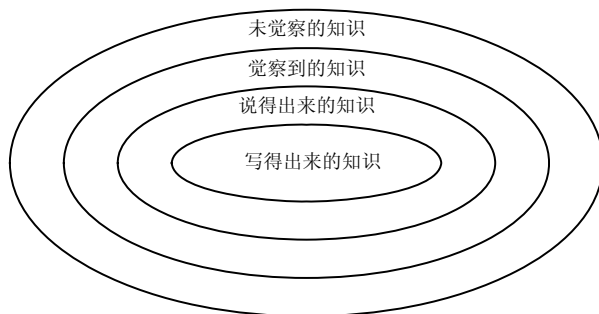


图 1.1 显性与隐性知识

从图 1.1 中可以看出,写得出来的知识仅仅是说得出来的知识的一部分,说得出来的知识仅仅是人所觉察出来的知识的一部分。在说得出来的知识之外的、人所觉察到的知识的那一部分,就是隐性(意会性)知识。实际上在隐性知识中还有一些连自己也还没有觉察到的不可言传的知识,在特定的场合下,却能发挥出来。这部分知识是通过无意识学习(或称不经意学习)得来的,因此有人称其为无意识知识。

隐性(意会性)知识来源于个人在生活实践过程中通过形体动作或感官接受而获得的感觉与体验,如人们掌握骑自行车和游泳就得靠亲身实践,过去手工作坊中师傅带徒弟进行像打铁、绣花等工艺操作,徒弟也是在实际劳动中观察、模仿、体验而获得技艺的。还有体育和舞蹈动作,这些都是与形体动作有关的。所有这些都很难仅用语言表达和传授。另外一些是人们在处理实际问题时,通过直觉和感悟而获得的。这类知识具有很强的个人特性,包括人的价值观和眼界以及洞察力和预见性,很难甚至根本不能通过

语言表达和传递。

隐性知识在人类获得知识的过程中起着极为重要的作用。但是关于它的研究，只是近年来由于它在创新与决策中的重要作用才引起人们的关注。

对于隐性知识的理解也是各不相同的。归纳起来，就有“不能用语言表达”、“不易用语言表达”、“不曾用语言表达”三种理解。

隐性知识中有一些经过转化而能够独立表达和传授的，就形成了显性知识。这包括人在工作中获得的经验和感受，但在转化过程中，一些富有个性化的因素也就遗失掉了。另外一些是不能转化的，这包括人的悟性、洞察力乃至价值取向，只有掌握这类知识的人才能亲自使用它。

把知识分为显性和隐性是一种二分法，对于建立概念是需要的，但是实际生活中的知识除了极少部分是纯粹显性（如一些定理定律）和纯粹隐性（如人的悟性）以外，大部分都是具有不同的显性度和隐性的，因此可以看作是像图 1.2 那样的、按知识的显性度排成的知识谱系中的某一部分。例如，有的知识一部分说得出来而另一部分说不出来，说得出来的知识一部分能够写出来而另一部分不能写出来，就是处于谱系的中间位置。

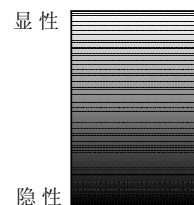


图 1.2 知识谱系

知识又可分为个人知识和组织（群组、企业）知识两类。

由于知识的产生来自人的实践与认识，知识是由个人产生的，离开个人，组织无法产生知识。但在社会活动中，组织也具有自己的知识，特别是表现为企业所掌握的技术、专利、生产和管理规程，有的已嵌入了产品与服务之中。组织知识是将个人产生的知识与其他人交流而形成并结晶于组织的知识网络之中的。个人只能获得和生成专门领域的知识，而在创新活动中，需要综合各种知识，需要转化为生产力，这就需要组织知识。

组织知识一般是指像企业、院所这样的组织所拥有的知识，在更大范围，还有一个国家、民族所积累的知识（当然还可以扩充到全人类的知识），这部分知识可以叫做社会知识。它反映了一个国家的文明程度和科技水平，是在不断地增加和发展的。其中既有本国的创造，也有从其他国家交流学习来的知识。

个人的知识一部分是从社会知识中通过学习获得的，也有一部分是对已有的知识和实践经验经过加工获得的成果。这部分知识经过语言、文字或实践活动可以转化为社会知识，但是知识不像实物那样转化后就失去原来的存在，而是仍旧保存在个人身上。

正像个人在社会环境中工作和生活时，不断积累经验、增长阅历，把它们保存在个人记忆中那样，组织在环境中生存也会积累经验，增长经历，把它们保存在所谓“组织记忆”中。这些集体的经验、阅历就是组织知识。对组织来说，下次遇到一个新问题，可以先去组织记忆中寻找过去在同样或类似情况下的成功解决经验，作为处理当前问题

的方法。如果没有，就在类似的经验基础上通过创新解决问题，从而创造出新的知识，加入记忆准备以后应用。

组织记忆包括文件、图纸和规程等，这些可以认为是组织的显性知识；此外还有组织的隐性知识，存在于个人记忆和已形成的人与人之间的默契之中。在企业的生产、经营活动中，需要把个人的知识协调起来才能使组织知识起作用。

人们不但关注个人的显性与隐性两类知识，而且关注组织的显性与隐性两类知识。特别是组织的隐性知识，那是买不来、偷不走、难以学到的，所以常被认为是企业的核心竞争力。

在人类的生产和生活中，除了“技术知识”外，还需要一种协调分工生产的“制度知识”，也就是关于组织管理的知识。正像工具是物化了的关于某种专门生产的知识一样，“制度”就是物化了的关于协调分工的知识。

知识又有经验知识和理论知识之分。经验知识乃是人们在长期的劳动和生活中通过感官体验获得的有使用价值的知识，包括各种手工技艺、服务经验、生活经验和人际交往经验等。经验是人类通过生产和生活，反复实践、逐渐感知和总结事物的形态与活动的技巧而获得的，大部分不是一学就会的，需要反复地慢慢体验。经验知识多半是隐性的，但有些可以转化为显性知识。

理论知识是通过实践获得的感性材料经过归纳整理、抽象而形成概念与公理，再进一步进行逻辑演绎而形成的假说和原理。它不是由人的感官直接获得的，而是通过人的大脑对客观现象的间接反映，寻找对这些现象的解释而获得的知识。

经验知识可以通过理性思维转化为理论知识，而理论知识应用于实践还得通过实践的检验和修正。在实际生活中，理论知识的应用经常伴随着经验知识；而相对应的是：显性知识的应用经常伴随着隐性知识。在复杂的人类实践和认识过程中，两者的界限常常是无法绝对划清的。

1.6 知识的作用与影响

知识对社会的作用与影响，是这一节所要探讨的内容。

1.6.1 知识的作用

知识的作用首先表现在生产力方面。我们不妨循着历史来考察生产力的发展。从古代的简单的手工劳动方式，经过机器与电力相结合的现代生产方式，正在走向以智能工具为主体的生产方式。生产力是由劳动资料、劳动对象和劳动力三者（三要素）结合构成的。上述三要素都是最基本的。知识对这三种要素起到了越来越显著的作用和影响。

在劳动资料、劳动工具方面，原始石斧的磨制成功，在如何选材、怎样磨削、选什么角度等方面，就已经开始积累经验，产生了知识。动力机械与工具机的发明与应用，更是科学技术知识的作用结果。工业经济时代的特点是用机器来制造机器，到现在，以计算机为基础的智能机器的出现和推广应用，使得新经济时代的特点是用机器来设计和控制机器，这都是知识的深入发展与应用的结果。

在劳动者方面，由于科学的普及与教育、培训的实施，从事劳动的人的知识水平逐步提高。劳动者不再单纯凭借简单的体力劳动来从事生产，而是把基于知识的脑力劳动与体力劳动相结合的新型劳动能力用于生产实践，操作机器。特别是在高新技术产业以及用高新技术改造过的现代化传统产业的生产过程中，生产者更多的是运用知识来监督和控制生产过程。

在劳动对象方面，由于科学技术的发展，人类不但可以更加有效地对已发现的自然资源加以使用，而且不断地利用知识发明和加工制造出新的材料，制造出新的产品。可以说劳动对象的范围也在不断地扩展。

人类在生产物质产品的同时，也生产了知识产品。因此可以说现代化的物质生产是在知识产品的指导下进行的。

从历史上来看知识的产生，它经历了一个从个体研究到集体协作而逐渐社会化的过程。社会对知识（这里主要说的是科学技术知识）的需求反映为市场对知识生产的推动与控制。

除此之外，知识还促进了产业结构的变化。从全球范围来看，工业化经历了三个阶段，每个阶段都有它的主导产业。早期的主导产业是纺织业和钢铁业，中期的主导产业是机械制造业和汽车业，后期的主导产业则是以微电子和软件为基础的计算机与通信相结合的信息业。随着知识的作用日益扩大，像软件、信息服务、专业科学研究以及咨询业在经济活动中的作用越来越重要，这一类知识产业正在形成主导产业，也就是下面将要讲到的创造力产业。产业结构的这种变化在很大程度上是受到科学技术知识发展的影响，这种变化将使得国民经济从资金密集型和能源密集型向技术密集型、知识密集型转化。

产业结构变化的另一种趋势是高新技术产品向传统产业渗透，促进了传统产业的现代化。例如，中国现在虽然还处在工业化的中期，但可以应用信息化来带动工业化，这与发达国家同一阶段的产业结构有所不同。另外这种渗透还催生了一批所谓“边沿产业”，如汽车电子产业。

知识的作用引起就业结构的变化。因为从事体力劳动的人数相对减少，而从事信息工作的人员（可称为信息工作者）数量相对增多。而信息工作者又可分为数据工作者（如统计员）和知识工作者（如工程师和软件开发人员）两类，其中知识工作者的比重会越来越大。

知识具有社会推动力，它在社会的各个领域，包括经济、社会、科技、文化诸多方面都时时刻刻发挥着作用，通过人的社会活动或者知识劳动表现出来。对个人来说，把积累的知识加以组织和应用，能够提高认识能力和实践能力，扩大劳动成果；对一个企业、院所、社区乃至一个国家来说，组织的知识体系能够提高组织集体的认识水平和实践能力，提高综合实力。这也就是为什么人们越来越关注知识的获取、传播、应用和创新的基本原因。

知识的推动力还具有社会公共性。一个人、一个组织或者一个地区要想长期垄断知识的推动力是比较困难的。因为知识是可以流动的，随着知识的流动，影响力也在转移。个人或者组织一般只能控制或垄断某一局部的知识，不能控制或垄断所有的知识。从历史上看，知识一旦产生，总会冲破重重阻力，流入社会的各个领域和各个角落，在流动过程中和为人使用的过程中表现出自身的推动力。

知识推动力在作用过程中显然也会遇到阻力。这种阻力来源于个人或集团的自身利益考虑。知识也有可能被滥用而产生知识的异化，如农药的滥用产生对人畜的危害和环境的污染。但是人类一旦掌握了有关负面影响的知識，就会逐渐克服这些异化现象。所以从长远来看，知识推动力是在促进人类文明的。

1.6.2 知识经济及其特点

1996年，经济与合作发展组织（OECD）发表了题为《以知识为基础的经济》的报告^[5]，在报告的一开始就提出：“以知识为基础的经济这个术语的出现，表明了人们对知识和技术在经济增长中的作用有了更充分的认识。知识，作为蕴涵在人（又称人力资本）和技术中的重要成分，向来都是经济发展的核心。但是，只是到了最近几年，正如知识的重要性在增长一样，人们对于知识重要性的认识也进一步深化。”

20世纪90年代后期，在这个报告的影响下，无论是在国外还是国内，都引起了研究讨论知识对经济影响的热潮。当时按照该报告的提法，人们使用的还是“以知识为基础的经济”一词，到了1997年2月，当时的美国总统克林顿在公开演讲中使用了“知识经济”的提法，后来的讨论就直接使用了这一名词。根据当时比较一致的理解，知识经济是建立在知识创新、传播和使用基础上的经济。

人们在研究讨论中认为，知识经济是继农业经济、工业经济之后的一个新的经济发展阶段，它将引起生产方式、生活方式、交往方式以至于思维方式的重大改革。

人们认识到，知识经济是一种新的经济形态。这种形态是在发达国家初步完成工业化以后，由于以信息技术为核心的高新技术及其产业的发展，传统产业的知识含量也同时得到提高，使知识密集型产业成为国民经济的支柱产业和经济增长的根源，借助于经济全球化而逐步形成的。

在有关知识经济的讨论中，人们也批判了一些片面的观点，并开始认识到：作为一种以知识为基础的经济形态，并不是说以前的其他经济形态都不以知识为基础；知识是重要的生产要素，但也不是说可以不以其他生产要素为基础。

实际上，知识作为一种可重用的社会资源，它的生产、分配和使用早就经济运行的基础之一。在游牧社会和农业社会阶段，人们已经利用以经验为主的知识；到了工业社会阶段，系统化的理论逐渐形成，人们利用的知识已带有科学色彩，但是其基础地位与作用却没有得到人们普遍的重视。只是近年来，科技进步对经济、社会生活的推动作用与日俱增，知识在经济生活中的基础地位开始为人们所密切关注。

在今天的经济活动中，土地、劳力、资本仍旧是重要的生产要素。特别是在发展中国家，人均物质产品的消费水平还很低，物质生产仍旧是经济成长的主要成分。随着科技对经济的贡献日益加大，产品、服务中知识含量的不断提高，知识将逐步成为促进经济增长的关键因素。知识作为一种生产要素，与其他生产要素相结合才能发挥作用。此外，知识对其他生产要素也是有积极影响的，如可以降低原材料的消耗和能耗，提高劳动生产率。

可以将知识经济理解为一种新的经济形态：从生产要素的角度来看，知识已经成为经济发展的最重要的要素，它不但在工业经济阶段起作用，而且在知识经济阶段，知识的作用更加显著，一方面要在改善资金和资源的利用方面起作用，另一方面又有物质要素所起不到的独特的作用。这在像软件这样的产业中表现得非常突出。

知识经济又可以理解为一种新的经济结构：从产业结构来看，高技术产业和知识密集的现代服务业逐步成为经济的主体。这类产业可以称为知识型产业。当前这类产业发展比较迅速，在各国的国民经济中的比重日益显著。

当然，这并不意味着农业和传统工业已经无足轻重。人们的衣食住行都还离不开农业和传统工业的产品，但是在知识经济阶段，农业和传统工业由于基于知识的高技术产业和现代服务业的渗透和影响，也已经现代化、知识化了。

知识经济还可以理解为一个新的发展阶段：到20世纪末，出现了以信息与知识的开发和生产为特征的新的发展阶段。在这个阶段，一些国家在物质产品方面的供给相对丰富，转而追求更高层次的发展，这时信息与知识成为发展的关键因素。原来对能源、物质资源的过分依赖和供应的紧张程度由于科学技术在节能和节约原材料方面的作用而得到缓解。原来人们从主要依靠体力劳动谋生发展到体脑并重，体力劳动和脑力劳动逐步融为一体，而潜在的智力因素逐步占有重要地位。

目前知识经济这种形态即使在发达国家也才是初见端倪，因此对于“知识经济”的概念也还只能是随着经济与科学技术发展和理论创新来逐步明确。根据当前的理解，这个新的经济阶段具有下列特点：

(1) 知识密集型的高技术产业比重越来越大。由于高科技的发展日益形成新的产业, 这些产业都建立在知识基础之上, 因此具有较高的经济效益。

(2) 从业人员结构发生了很大变化。更多的劳动者转移到从事信息和知识处理的岗位上去。同时对在一线工作的新型工农业劳动者的知识水平要求也提高了, 成为知识工人。

(3) 资源环境发生了很大的变化。农业经济时代土地和劳动力是生产的基本要素; 工业经济时代资本、原材料、能源等是重要的生产要素, 是以物质为基础的; 而到了知识经济阶段, 同样的物质资源, 通过知识的作用, 可以得到更多的产出, 知识和信息自然也成为生产的基本要素, 知识已经从非独立因素发展成为独立因素, 由潜在生产力发展成为现实生产力。

(4) 资产投入的无形化。知识已经成为一种无形的资本^[9, 10]。对知识密集型企业, 投入的不仅是像资金、设备等有形的资本, 还需要知识、智力等无形资本的投入, 有知识的个人或者组织常常用自己的知识入股。

(5) 研究与开发的投入大量增长。创新成为发展的关键。由于现在产品更新换代越来越快, 市场上的竞争也越演越烈, 因此研究与开发成为企业生存与发展的决定性因素, 产品、流程、经营管理模式的创新成为可持续发展的关键因素。

(6) 学习有着特别重要的意义。由于知识的更新在加快, 人类的知识正在呈爆炸性增长。无论是个人还是团队, 为了获取知识, 都需要不断地学习。因此知识经济也可以认为是学习经济。对一个组织来说, 面临着把自身转变为一个学习型组织的任务, 只有学习才能了解不断变化着的形势和所处的环境, 应对发展做出正确的判断和决策。对创新过程来说, 更需要经常学习。

(7) 教育成为经济发展的基础之一。由于知识的获取、处理和集成都是由人来做的, 知识的创造更是由人来完成的, 都是以人为中心的, 因此对人的素质提出了更高的要求。这就需要大力发展和提高教育事业, 培养一代富有创造性的新人。

(8) 人与自然协调、可持续发展。由于知识密集型产业所耗用的能源与其他资源相比都有所降低, 因此有利于可持续发展, 使得人与自然和谐相处。

科技进步促进了经济的发展, 而经济反过来又推动了科技进步, 两者相互作用与融合, 促成知识经济的产生。知识已经从书斋和实验室走入经济生活, 成为重要的经济资源、生产要素和资本。

1.6.3 创造力经济及其影响

2008年, 联合国所属的贸发会(UNCTAD)、开发署(UNDP)、教科文组织(UNESCO)、国际贸易中心(ITC)的专家联合发表了一个名为《创造力经济》的长篇报告, 对创造

力经济和创造力产业都作了详尽探讨^[1]。创造力经济(Creative Economy)实质上就是以创造力为主要驱动力而形成的经济形态,强调经济增长与发展的驱动力是人的创造力。创造力是人所独有的生成某种新事物的能力,这种能力产生于人的大脑。人们产生了某种新的意念、主意和灵感,人们按照这些新的思想去探索、研究、开发,成为自觉的、有意识的“创造性活动”,“创造性实践”,取得“创造性成果”,这就是科学发现与技术发明的过程。人们也可以去调查、研究、实践和创作,形成文学、艺术、音乐和电影等作品或者社会科学的新理念、新观点或新方法。我们称其为“创造性思想观点”,或者“创意”。当然创造力活动并不完全与经济挂钩,如普通教育和公共事业中的创新活动,而与经济挂钩的部分就属于创造力经济。

创造力经济是知识经济的核心,创造力产业是创造力经济的物质基础。所谓创造力产业(Creative Industry),是指那些运用个人创造力和才华创造财富和就业潜力的行业。一般来说,其中一些是涉及科技研究开发的,如工业设计和软件,一些是涉及建筑设计和工程勘察设计;另一些是涉及文化传媒,如文艺创作表演、广播电视、动漫和电影制作的;还有的是涉及咨询策划,如市场调研、证券咨询和会展服务。其中有一些属于通常所说的文化产业。而在文化界则喜欢把创造力产业称为创意产业。所有这些产业的产出都是非物质产品,尽管其中的一些依附在实物之上,如音乐戏剧录在光盘上,但真正的价值不在物质载体本身。创造力产业之所以能形成产业,就在于这些精神产品能够直接进入市场。

到现在为止,对于创造力产业的界定还没有一个众所公认的看法,新的创造力产业也还在不断出现。创造力产业的兴起,一方面是受到经济的发展和城市转型的影响。城市在工业化向后工业化转变的过程中,经济结构从以生产为中心向以服务为中心转化,为创造力产业发展提供了资源和环境条件。另一方面是人类产生了大量新的需求,他们在追求物质享受的同时,还在追求精神满足,产生了对优秀文化产品的需求,新的感觉和体验成了幸福的主要源泉之一。在不同经济发展程度的国家,或者同一个国家的不同区域、不同人群,人们的需求存在着巨大的差异,为满足个性化的需求,要求不断有新的物质和精神产品出现。再加上科学技术的发展为创造力产业提供了大量的技术手段,特别是互联网的发展提供了一个巨大的创新空间,使得创造力产业的发展异常迅猛,尤其是在一些发达国家,创造力产业正在以比实物生产快得多的速度在增长。

创造力经济可以创造知识与知识产权,经过产业创新过程,逐步形成了大规模的知识经济。从广义上来说,创造力经济与知识经济在很广泛的范围内是互相融合的。

知识经济和创造力经济对经济、社会、生态和文化的影响表现在下面这些方面:

(1) 对产业结构有所影响,高技术与全球化是其特征。前面已经提到,由于高科技

的发展不断涌现出新的产业群,如信息技术产业群,生命技术产业群,新能源产业群,新材料产业群,核技术产业群,空间技术产业群,特别是文化产业群,等等。传统的制造技术也依靠高技术而进入以服务型制造为特征的新的发展阶段。这就使产业结构发生了很大的变化。

正因为知识经济和创造力经济需要快速、灵活和规模适度的运转方式,要求不断更新和转移技术,使各种资源在全球范围内合理配置,因此希望全球的经济成为一个统一的有机体,以便最大限度地发挥知识对经济和社会文化的决定性作用。而信息技术和高效运输技术的发展又为全球化提供了有效的工具。因此近年来形成了一个全球化的经济环境。

(2) 对投资模式有所影响,导致对无形资产的大规模投资。知识经济和创造力经济虽然需要资金的投入,但更需要知识、信息、智力等无形资本的投入,这对高技术产业和文化产业说来表现得尤为突出。目前发达国家许多高技术企业的无形资产已超过总资产的 60%。而无形资产的升值,也使社会的价值观发生了变化,拥有更多知识的人能够获得更高的报酬。

(3) 对企业活动有所影响。在工业经济时期,其重点在生产,而在知识经济阶段的重点则着重于生产阶段之前的研究开发和设计,以及生产阶段之后的市场开拓和营销。这两头的收益都要比生产收益高,因此许多发达国家的企业都把产品的加工转移到劳动成本低廉的发展中国家,而在本国只抓两头有高收益的环节。

(4) 对教育有较大的影响,更有利于人的全面发展。在知识经济和创造力经济阶段,不但正规的全日制教育得到高度发展,而且由于劳动者需要不断补充和扩展自己的知识,因此各种继续教育、培训方式和机构也有所发展,因为这关系到知识的传播和再生产。知识经济和创造力经济需要充分开发人力资源,需要全面发展的新人,而新经济创造的物质条件也有利于人的全面发展。

(5) 对人的劳动方式和生活方式也有所影响。在知识经济和创造力经济阶段,从事体力劳动的人数逐渐减少,而从事脑力劳动的人数不断增加。由于机械化、自动化、智能化的发展,使得体力劳动的强度逐步降低,劳动环境逐步改善。从事体力劳动的人的知识水平也在逐步提高。更多的劳动者直接从事的工作,将从以物质产品生产为主,逐步发展到以信息和知识资源的开发为主。正是因为许多智力劳动方式具有个性化和弹性化的特征,因此可以使人们相对自由地支配时间,这让人工作与生活方式也产生了相应的变化。此外,由于工业经济高度的分工,使得每个劳动者只掌握极为简单的技能,而知识经济和创造力经济却能使劳动者掌握手脑并用的多种能力,这将有利于人的全面发展。

(6) 对人与自然的协调发展产生良性影响。工业经济在创造了巨大物质财富的同时,消耗了大量不可再生资源 and 能源,导致了环境的恶化。目前,人类正在受着温室效应、酸雨、土地沙漠化、海洋赤潮等环境问题的困扰,不可再生资源所能继续开采和使用的时间也是有限的。人与自然的这些冲突是工业文明造成的后果,不可能在工业经济的框架内解决。知识经济和创造力经济注意到了对资源和能源的节约和对环境的保护和恢复,为人和自然的协调发展开辟了道路。

(7) 对政府的功能有所影响。从国家层面来说,知识经济和创造力经济关系到如何建立国家创新体系,营造创新环境以及以知识为重点的综合国力不断增长和提高的问题;而在城市层面,关系到如何建立知识城、创新城的问题。



思考与讨论题

- (1) 试举出天然信息、人工自然信息、社会信息、思维信息的一些实例。
- (2) 试举出一些对你的学习和生活最具影响的信息的例子。
- (3) 用你身边的例子来说明信息化是在不断发展的。
- (4) 你能举出一些隐性知识的例子吗? 并说明其中哪些是可以转化为显性知识的? 哪些不能?
- (5) 有人说书本上的内容不是知识,仅仅是信息,你对这个观点有什么看法?
- (6) 请你思考一下工业经济与知识经济的差别以及两者之间的关系。



参考文献

- [1] 钟义信. 信息科学原理[M]. 第3版. 北京: 北京邮电大学出版社, 2002.
- [2] 钟义信(主编). 信息科学与技术导论[M]. 北京: 北京邮电大学出版社, 2007.
- [3] Wiener N. Cybernetics [M]. 2nd ed. New York: MIT Press, John Wiley & Sons Inc., 1961.
- [4] 王众托. 企业信息化与管理变革[M]. 北京: 中国人民大学出版社, 2001.
- [5] 经济合作与发展组织(OECD). 以知识为基础的经济[M]. 北京: 机械工业出版社, 1997.
- [6] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [7] 柯平. 知识管理学[M]. 北京: 科学出版社, 2007.
- [8] Polanyi M. The Tacit Dimensions [M]. New York: Doubleday Anchor, 1966.
- [9] Steward T. Intellectual Capital [M]. New York: Doubleday/Currency, 1997.

- [10] Liebowitz J. Knowledge Management Handbook [M]. London: CRC Press, 1999.
- [11] United Nations UNCTAD. Creative Economy [R]. 2008.
- [12] Nonaka I, Takeuchi H. The knowledge creating company[M]. New York: Oxford University Press, 1995.
- [13] Davenport T. Information Ecology [M]. New York: Oxford University Press Inc., 1997.
- [14] Simon H. The Sciences of Artificial [M]. Cambridge: The MIT Press 1996.

第 2 章

信息管理概述

内容提要

本章是对信息管理的概括介绍。首先介绍有关信息管理的含义，其次介绍不同层次的信息管理过程。然后从信息资源管理、信息系统管理、信息组织与人员管理方面阐述信息管理的内容，并介绍了信息服务的内容。最后介绍信息管理中人的因素问题。

本章重点

- 信息管理的定义与特征
- 信息过程
- 信息管理的内容
- 信息管理中人的因素



2.1 信息管理的含义

我们先从信息管理的概念、层次和它的职能开始介绍信息管理的概况。

2.1.1 信息管理的概念

第1章曾经讲到,随着信息作用的日益显著,人们的信息活动日益频繁,需要有目的、有计划地对信息进行管理。其实人类对信息进行管理的历史是比较久远的。古代从以图画方式记录和传递信息,通过象形文字发展到正式文字的产生,提供了越过时空进行信息记录和传递的工具。当时对文书档案的存储整理与通过驿站和烽火台对信息的传递,都是信息管理的雏形。到了工业革命阶段,印刷术的应用提高了信息复制和汇集的效率,交通工具的进步密切了各地的联系,加强了文献的传播。特别是电工技术中的电信技术的发展,为信息的采集和交流提供了新的手段。这时文献资料、新闻报道的组织与传播已经有专门人员和组织(如现代化的图书馆、信息情报中心和电信部门)来进行,管理思想和手段也不断地发展和进步。等到20世纪的中叶,信息技术迅猛发展,特别是现代化的通信技术和计算机技术的发展以及两种技术的有机结合,使得信息的获取、传播、处理和利用有了飞跃的进步,而社会经济的发展对信息的需求无论是数量上还是质量上都有更高的要求,面对海量信息亟需加以组织管理,因而把信息管理推进到一个新阶段。

信息管理现在也还没有众所公认的定义,因为着眼点不同,就会有不同的定义方法。比较普遍一点的看法,是把信息管理定义为:人类为了有效地开发和利用信息资源,以现代信息技术为手段,对信息资源实施计划、组织、指挥、控制与协调的社会活动^[1]。

这个定义涉及信息管理的三个要素:人员、技术和组织。涉及所要管理的两个方面:信息资源和信息活动。管理活动的基本内容和其他管理活动一样,是计划、组织、指挥、控制和协调。

为了深入了解信息管理,我们需要了解它的社会功能。信息管理的社会功能有下列三个方面:

- (1) 开发信息资源,提供信息服务。
- (2) 合理配置信息资源,满足社会对于信息的需要。
- (3) 推动信息产业的发展,促进社会信息化水平的提高。

从这些社会功能来看,信息管理的对象集中在信息资源和信息活动两大方面。

信息资源是包括信息内容、信息技术和信息人员这几类要素的有机集合。其中信息

内容是信息资源的核心要素，它以文字、声音、图形和图像等记录符号描述与表达事物的运动状态与变化形式，并以消息、资料 and 知识等形式被人们处理和利用，或者以数据、文献或某种实物的形式被记录、传递和储存。信息技术作为技术手段，对信息的获取、传递、存储、处理和利用提供条件，克服了人类自身的能力局限，扩大了处理信息的能力。正是信息技术的发展和广泛应用，才使信息本身得以发挥作用。信息人员则是控制组织与信息资源、协调信息活动的主体，一切信息活动都是按照信息人员的意图进行的。这三种要素相互联系相互作用，共同构成具有同一功能的有机整体——信息系统。通常人们对于信息资源和信息系统的理解，常常存在见物不见人的片面性，其实“人”才是资源中的主体要素，也就是信息系统中最为活跃的要素。

信息活动是指人类围绕着信息资源的形成、传递、处理和应用开展的服务活动和管理活动。信息活动按照过程来看，可以分成两个阶段：

(1) 信息形成阶段。这个阶段包括信息的产生、采集、传播、处理和存储。其目的在于形成可以利用的信息。

(2) 信息开发利用阶段。这个阶段包括信息的检索、传播、吸收、分析、选择、评价与利用。其目的在于使信息发挥作用。

2.1.2 信息管理的层次

信息管理是在不同的层次上进行的。从大的方面来划分，有宏观管理和微观管理两个层次。但是如果再细分一下，可以分为国家层次、组织层次和个人层次三个层次，此外还存在一个中间层次，即社会层次。下面依次研究各层信息管理的任务和内容。

1. 个人层次的信息管理

个人层次的信息管理实际上就是由个人来管理和他个人有关的信息，这种管理纯粹是个人行为。其实人类很早就自觉不自觉地管理个人的信息了。个人的信息处理与信息管理的融为一体。例如，他不断地获取与他有密切关系的信息，并以各种方式加以存储（记忆在脑子里和书写在介质上）。同时不断地通过思维加以处理和利用。这方面的问题涉及人的认知和心理能力。随着人的活动范围的不断扩大，个人管理的信息不只是与他本人直接有关的信息，还有与自然环境和社会环境有关的信息。但是在目前这方面的研究还不多，还不是信息管理中的主流和重点，相信日后会越来越多起来。

2. 组织层次的信息管理

组织层次的信息管理是像企业、学校、研究机构、政府机关等组织为了实现自己的战略目标，使用自己的各种资源，对所需要的信息进行获取、处理和利用进行的管理。组织内部常常成立专门的信息管理机构（如信息中心、信息部），成为一个新兴的独立职

能部门，来专门从事信息的处理和管理。最高层领导会有一位副职专门管理信息工作，还有专门的信息主管（CIO，这在后面还会讲到）。这一层次的工作重点，首先是做好组织内部的信息资源和信息系统的管理。但是随着社会经济与科学技术的发展，一个组织与外界的联系日益增多，而且联系方式也日益多样化，因此组织的信息管理还要考虑与外界信息组织共同工作的问题。本单位的信息设施再完善、信息资源再多，也无法满足日益增长的需求，还需要仰仗组织外部的专门机构，因此就出现了社会化的专业信息公司和咨询机构等。这些公司和机构本身的信息管理又各具特色，因而成了组织层次信息管理的新问题。

3. 国家层次的信息管理

国家层次的信息管理针对的是整个国家的信息环境。为了维护本国的信息秩序，确保社会信息活动能够向有利于国家利益的方向发展，国家的信息管理从宏观上对全国的信息资源、信息设施和信息活动加以控制，其目的在于保证各个层次的信息活动都能够为全体人民的工作和生活带来方便和效益。

国家的信息管理和政府的信息管理还是有差别的。国家层次的信息管理是对整个国家的信息环境进行的一种规范控制，是宏观管理，而政府的信息管理一方面是对宏观环境施加影响，保证国家层次的信息管理得以实现；另一方面是政府为了进行政务工作而进行的信息活动的管理。

国家层次的信息管理采用的手段有如下三种：

（1）经济手段——通过税收、价格、利率、投资等经济杠杆，引导社会信息活动向健康方向发展。

（2）行政手段——通过制定各种政策，如产业政策、信息政策、贸易政策和各种规划、计划、各种命令等，直接进行控制和管制。

（3）法律手段——由国家最高权力机关以立法名义制定各种法律，从而进行控制。

对于像中国这样幅员广阔的国家，为了对一个特定地区进行信息管理，也可仿照国家层次进行引导和控制，但这仅仅是国家统一管理的一个分支。

4. 社会层次的信息管理

社会层次的信息管理是面向社会公众提供各种信息服务的机构所进行的信息管理，如图书馆、档案馆和信息情报机构等，这多半是非营利性的公益机构。这与第二类——组织层次的信息管理有一定的类似，但它们是面向全社会的。

在上述几个层次中间，以组织层次的信息管理最为典型，后面的讨论就专门注重于这个层次。

2.1.3 信息管理的职能

本书在前面的信息管理定义中曾经提到,信息管理是对信息资源实施计划、组织、指挥、控制与协调的社会活动。这些活动就是信息管理的职能。下面分别对各项职能加以分析。

1. 信息管理的计划职能

从管理的一般原理来看,计划职能包括确定目标、制定总体战略和分计划,协调组织的活动。信息管理的计划功能,是围绕着信息活动的整个管理过程,通过对组织内外的调查研究,并对未来进行预测,根据组织整体战略确定信息工作的总体目标,然后分解出子目标和各阶段的任务,制定各种管理计划,作为行动指南。

信息管理的计划主要包括信息资源计划和信息系统建设计划两项。信息资源计划的内容是信息的采集、加工、存储、利用和更新维护等方面的具体计划。信息系统建设计划是关于信息系统建设的行动安排,包括信息系统建设的工作范围和对人、财、物和信息资源的要求,成本估算和工作进度安排等。

2. 信息管理的组织职能

一般说来,组织的职能在于决定本单位需要完成什么任务,怎样完成,由谁来完成。信息管理的组织职能包括信息系统的研究开发及其管理、信息系统的运行维护及其管理、信息资源管理与服务、信息管理组织效能的提高。

信息管理的组织需要对内提高信息系统研究开发和组建的效率与质量,对外提供优质的信息服务、技术支持和人员培训。

3. 信息管理的指挥(领导)职能

从普遍意义上来说,领导的职能在于采取有效的方式指导和激励所有的参与者,并且解决矛盾和冲突。信息管理的领导职能,在于信息管理的领导者对组织内部所有成员(或者某一信息部门的领导者对部属)的信息行为进行领导、指导或施加影响,使他们能够自觉、主动地为了实现既定目标而发挥潜力。

具体的工作包括参加高层决策,提供决策所需要的信息,提出建议;负责组织信息系统的开发和运行维护;推广新的信息方法和工具;对外加强联系与协作,等等。

4. 信息管理的控制与协调职能

从普遍意义上来说,控制职能包括对计划的实施和组织的活动进行监督和调控。信息管理的控制与协调职能是为了确保信息管理的目标和计划能够顺利实现。

具体的工作包括对信息工作不断进行监测和评价,发生偏差时及时纠正。此外还要在计划执行过程中,当内外环境发生变化时及时修订计划或者制订新的计划,并进行部

署。这些工作都是需要经常进行的。

上面所说的主要是针对像企业或者院（所）这样的组织的信息管理职能。至于大到一个国家、一个地区或者一个大型集团，其信息管理的职能基本上也是这些，只是范围和权限不同，影响不同。

2.2 信息管理过程

本书前面曾经讲到，在信息活动中，信息要经过生成、采集、传播、处理和存储、检索、传播、吸收、分析、选择、评价和应用等过程，这些过程有些是由人来操作的，有些是用信息工具完成的。一些基本操作可以称作信息的处理，而对这些操作的组织管理就是信息管理工作。最初由于信息量还不那么多，多半是由个人来完成，所以信息处理和信息管理难以明确划分，这正像在小作坊中，生产的操作（加工）和管理（原料和工具的管理）难以明确区分是一样的。

等到信息量大起来，处理的方式和工具多起来以后，信息管理就独立出来，而管理的层次也多起来，管理的内容和方法也更加丰富。在信息管理领域，有许多学者曾经把所有的信息管理过程分门别类地列举出来，这种分类列举有助于人们对信息管理的动态行为有个系统的认识。

下面举出一种按照战略、战术和运作三个层次，对不同管理职能中的信息管理过程进行划分的情况。这种划分所包括的过程比较适合像企业、研究单位、管理机构等组织的情况。

2.2.1 战略层次的信息管理过程

战略层次的信息管理过程包括战略规划和战略控制两个方面。

1. 战略规划方面

战略规划方面包括：

- (1) 经营战略规划过程。这是在最高层次定义组织的整体信息战略。
- (2) 体系结构巡视与定义过程。在信息战略明确后，定义组织的数据体系结构、信息体系结构、知识体系结构、应用体系结构，以及组织的信息技术体系结构，然后将这些体系结构加以集成。

2. 战略控制方面

战略控制方面包括：

- (1) 信息战略规划过程。为了达到企业信息战略目标确定总方向。
- (2) 信息战略控制过程。对战略的执行进行控制。

2.2.2 战术层次的信息管理过程

战术层次的信息管理过程包括开发计划、资源计划、服务计划和管理计划四个方面。

1. 开发计划方面

开发计划方面包括：

- (1) 应用计划过程。这是在一段时期内要建立和修订的应用内容的集成。
- (2) 数据计划过程。回顾应用计划，确定数据需求。
- (3) 网络计划过程。确定组织的网络联通需求。
- (4) 系统计划过程。将企业的信息战略目标和方向通过一个系统的规划提出硬件、软件、网络以及其他设施的计划。
- (5) 项目计划过程。将所有的计划转变成可管理的实施项目。

2. 资源计划方面

资源计划方面包括：

- (1) 容量计划与管理过程。确定当前和未来所需资源，并加以管理。
- (2) 技术计划与管理过程。确定所需的人力资源，并加以管理。
- (3) 预算计划与价值管理过程。将各种计划的需求转化为财务要求，提出经费预算。
- (4) 供应商计划与管理。提出供应计划与外包的项目。

3. 服务计划方面

服务计划方面包括：

- (1) 服务层次计划与管理过程。通过协商确定服务的层次，预计服务的工作量，并加以管理。
- (2) 复原计划与管理过程。保证在出现技术故障和突发事件时能够保障系统连续工作，并加以管理。
- (3) 安全计划与管理过程。制订全面满足安全要求的计划，并加以管理。
- (4) 审计计划与管理过程。制订审计计划并加以管理。

4. 管理计划方面

管理计划方面包括：

- (1) 管理系统计划过程。通过项目管理方法，管理上述计划的实施。
- (2) 管理系统监测过程。对实施进行监测。

2.2.3 运作层次的信息管理过程

动作层次的信息管理过程包括项目管理、开发与维护、资源管理、信息服务和服务控制五个方面。

1. 项目管理方面

项目管理方面包括：

- (1) 项目指派过程。确定项目的范围、领导方式以及与合作伙伴的关系。
- (2) 项目调度过程。详细规定项目的目标、资源、时间进度、任务划分、组织以及交付形式。
- (3) 项目控制过程。对项目各阶段进行监测，解决产生的问题，提出进度报告。
- (4) 项目需求控制过程。在项目计划需要改变或调整时，同意或者拒绝变化的请求。
- (5) 项目评价过程。在项目完成后确定是否所有需要交付的均已完成，并编写文档。

2. 开发与维护方面

开发与维护方面包括：

- (1) 软件开发与更新过程。开发与更新操作系统、应用软件以及其他支撑软件。这需要详细定义需求，进行内部和外部设计，将数据结构化，与用户对设计进行协商，开发并测试程序，最后进行集成与安装。
- (2) 软件采购与更新过程。对操作系统、应用软件以及其他支撑软件进行采购与更新。一般是先对理想的需求加以定义，对供应商提供的产品检查其一致性，并与其反复商谈，最后进行采购、软件安装和编写使用说明书。
- (3) 硬件采购与更新过程。这是选择、安装和更新信息系统硬件设施的过程。也是先要详细确定需求，然后选定设备，制订安装布局，对设备进行测试然后安装。
- (4) 系统维护过程。维护并纠正硬/软件中的缺陷。
- (5) 系统的协调与平衡过程。通过定期测试保证在有新的变化时系统能够正常运行。

3. 资源管理方面

资源管理方面包括：

- (1) 变更控制过程。监测、发现、评估信息系统与工作过程中的变化，采取措施，使其尽可能小地影响系统工作并降低风险。
- (2) 资产控制过程。对所有的信息资源及其库存进行监控。

4. 信息服务方面

信息服务方面包括：

- (1) 信息生成过程。对工作岗位、文件流转和数据进行调度,以满足服务要求。
- (2) 服务销售过程。发现未来市场需求,销售信息服务。

5. 服务控制方面

服务控制方面包括:

- (1) 生成与分配的调度过程。将服务协议变成工作计划。
- (2) 问题控制过程。发现问题,寻找事先确定的解决步骤并解决问题。
- (3) 服务评价过程。将实际绩效与协议书上的规定进行比较,进行评价,对变动进行说明。

2.3 信息资源管理

这一节讲的是信息资源及其管理。

2.3.1 对信息资源的进一步认识

在本章的 2.1 节曾经提到信息资源,目前对于什么是信息资源有着不同的认识 and 理解。其中主要的有广义与狭义两种理解。

1. 广义信息资源

像在 2.1 节所说的,信息资源包括信息内容、信息技术和信息人员。这是广义的理解,指的是信息及其相关因素的集合。

2. 狭义信息资源

狭义信息资源仅指信息内容(如果对它再细分的话,还可分出信息内容本身和信息载体,因为信息内容总要依附在载体之上)。

要想使信息发挥作用,广义信息资源的各要素需要组织成有机的整体,这就是信息系统。在系统中,信息人员是信息的生产者或加工者,不但生产原始信息,还要对原始信息加工或者再生产。信息技术则包括所有的对信息获取、信息存储、信息加工、信息传播和信息应用的各种设备(如通信网络设备、计算机设备、常规信息处理与存储设备等)、方法、计算机软件等工具,以及运行的制度和规程。以这样一个硬/软结合而以人为主的系统,对信息活动进行支持和保障。

现在信息资源已经和物质资源、能量资源并列,成为人类活动的三大资源。但是人类对信息资源的认识比较晚,这是由于历史上长期以来人类的个体分散活动所需要的信息量较少,信息又比较抽象,不像物质和能源那样能够看得见、摸得着。直到社会活动的规模越来越大,产生的信息量迅速增加,人们为了在海量的信息中有效地得到有用的信息,产生了对信息资源加以组织管理的需求。再加上信息技术的发展,使信息资源的

管理成为可能，信息资源管理的理念、方法和工具才发展起来。

如果从广义信息资源管理的内容来看，它和信息管理的含义是一样的，因此在许多文献中，使用信息管理这个名词，而在另一些文献中则使用信息资源管理这个名词。中国国内大多使用信息管理一词，而在国外则喜欢使用信息资源管理一词。

如前所述，在广义信息资源各要素中，信息内容是核心要素。关于信息系统管理和信息组织与人员的管理，本章后面都有专门的一节来探讨，下面先就信息内容管理来做一些研究。

2.3.2 信息内容管理

信息内容管理涉及信息的生命周期。信息不是生物，仅就它本身来说，没有所谓生与死的问题。但是从信息使用者的视角来看，从希望信息产生到使它发挥作用，最后到不再有用而废弃它，也和一个生命过程一样。因此可以把信息的生命周期看作是信息对主体所具有价值的存续期。整个生命周期可以分成几个阶段。对信息过程的管理是沿着整个生命周期进行的，而每个阶段内则要对信息内容进行运作。信息内容管理则是对这些运作的管理。它们与信息过程是融为一体的。

对每一个信息内容来说，都要经历下面五个阶段，可以考察各个阶段需要对信息内容进行的运作。

1. 信息需求的确定

信息需求的确定是催生信息内容的起点，这里要确定的是究竟谁需要某一信息内容，具体的信息内容是什么，何时何地需要，来源在哪里。

2. 信息的生成或采集

在确定了需求之后，就要确定如何生成所需信息，由谁来生成；或者采集所需信息，由谁在哪里（组织内部还是外部）采集。

3. 信息的处理

所谓信息处理，包括进行格式转换，建立索引，进行必要的重组或转化，此外还需要进行信息内容的存储。

4. 信息的利用

信息的利用指对已经加工过、可资利用的准确完整的信息，及时可靠地传递给使用者。

5. 信息的处置

在超过信息内容的有效期或者已经用过之后，这些信息内容就要处置掉，或者转到其他地方供以后再用。

在上述运作中，信息处理与信息内容管理是紧密结合的，有时候是融为一体的。

2.4 信息系统管理

信息管理是要通过信息系统来实现的，下面先从介绍信息系统开始。

2.4.1 信息系统的概念

信息系统乃是为了采集、存储、加工和传递信息以提供信息服务的有机整体^[2, 3]。一般的理解主要是指由计算机和网络以及软件所组成的技术系统，有时也称其为信息基础设施。其实信息系统不仅包括由计算机网络组成的现代化系统，还包括常规的通信系统（如固定电话与移动电话，邮递系统及公文人工递送的系统等）；不仅包括工具系统，还包括人际网络以及保障系统运行的工作规程等无形的要素。

信息系统的输入是通过各种方式采集来的原始信息，经过处理和传递，输出的是在内容和形式上都符合使用者所需要的信息。在这个过程中，计算机网络系统只是在做信息处理工作，而信息的输入和输出信息的利用还是靠人来完成，这也就是为什么要强调指出信息系统是一个由技术工具和人共同组成、而以人为主导的人-机系统的缘故。

信息系统中计算机的作用是为了对数据进行高速处理，网络的作用是为了把一个组织中分散在各处的信息集中起来，便于汇总和宏观分析。

信息系统的功能包括下列五种功能。

1. 采集信息

采集信息是信息系统运行的起点。一般分为原始信息采集和二次信息采集两类。原始信息采集是在信息发生的当时当地，从发生信息的实体中采集，并利用技术手段把它记录下来。例如，超市中从商品的条码中采集信息，在货物的电子标签上采集信息。目前在交通管制中和建筑物中装设的摄像头采集视频信息也是这一类。一般要求原始信息采集要做到准确、完整、及时。二次信息采集则是把已经脱离原来信息生成的实体而记录在某种介质上的信息进行采集。一般都是从另外的信息系统或信息记录、存储设备中提取。

2. 存储信息

采集的信息不见得都是即刻使用，为了留待今后处理或使用，需要将信息存储起来。信息的存储要进行分类、整理、保存、维护和更新，而在需要时又能及时迅速地取出。信息的存储要做到不丢失、不走样、不外泄，随时可以按照需要调出。

3. 加工信息

信息的使用除了一些极简单的场合（如数据查询）可以在采集后直接应用之外，大

多数情况下需要进行一定的加工处理,使其成为对主体来说直接可用的信息。对不同的用途,加工信息深度也不一样,有的仅仅是汇总、比较,有的则需要进行数值或逻辑的运算。这里所说的加工,不仅是使用计算机来处理,还包括人对信息的选择和判断。对于信息加工,要求切合实际用途,准确和及时。

4. 传递信息

当使用信息的组织有一定的规模,在地理上有一定的跨度时,要想将信息送达各处,系统必须有信息传递功能。此外,组织和个人还需要和外界进行联系,也需要传递信息。正如在前面说过的,信息传递是克服了空间障碍,信息存储是克服了时间障碍,是超越时间的传递。对于信息传递,要求及时、迅速、准确而不失真。

5. 提供信息

信息系统的最终目的是向使用者提供信息,这就需要在系统和使用者之间建立接口或者说界面。界面的作用是使用户能够按照他所习惯的方式获得所需要的信息。界面的另一个作用是提供使用者与系统进行对话。因此要求接口具有很好的用户友好性。

为了实现上面五种功能。信息系统都具备相应的子系统。这是任何类型、任何用途的信息系统都应该具备的。

最早出现的信息系统是为单一目的使用(如记账和统计)的数据处理系统(DPS),后来应用的领域扩展了,出现了所谓管理信息系统(MIS),不仅对数据进行登记和统计,而且进行一定的分析。此外还有办公自动化(OA)系统。再进一步出现了决策支持系统(DSS)以及主管信息系统(EIS)等。现在又逐渐演化像企业资源计划(ERP)、供应链管理(SCM)、客户关系管理(CRM)等系统。信息系统的分类也是多样的,有关这些系统由于有专门的课程讲述,这里就不详细展开了。

2.4.2 信息系统的体系结构与技术平台

从系統工程的角度来看,任何系統都有一定的功能和相应的结构。信息系统为了实现上面提到的各种功能,一定具有相应的结构。我们先从研究系统的体系结构开始。

系统的体系结构(Architecture)是用来表述系统某一方面特点的结构方式和模块组成的。它侧重原则、侧重方法而不具体规定技术与业务细节^[4],而更着眼于为实现系统功能所做的原则性的安排。这种安排着眼于各部分的关系应该怎样处理,使得系统在整个体系上结构合理,有关部分各得其所。

体系结构这个名词,其概念先是从一些工程系统中产生,最早是针对房屋、楼宇的整体布局的安排,使得各部分的空间位置合理匀称。后来这一概念推广到其他工程领域,就有了如计算机的体系结构等概念。在建立一个系统时,先得从全局的高度来考虑系统

究竟应该如何构成,以满足要求,也就是说,先要搭建系统的架构。系统的体系结构正是从不同的视角来描述系统架构的。

这里举出一个组织(如一个企业)的信息系统体系结构的例子,其中包括业务体系结构、应用体系结构、信息体系结构和设备体系结构四种。

1. 业务体系结构

业务体系结构,描绘的是企业在经营管理中具有的功能范围与业务活动过程。功能范围是指企业中主要的业务领域,如计划、财务、物资、生产、销售、分配和人事,等等。业务活动过程是实现各部门功能以及部门之间联系的活动集合,一个过程可能包括许多项活动。它们是由功能范围中划分出来的功能经过分解而产生的。

2. 应用体系结构

应用体系结构要考虑:

- (1) 在哪些环节上要用到信息技术;
- (2) 是什么人在应用它;
- (3) 完成什么功能;
- (4) 人与系统的分工和衔接。

这里实际上是把业务流程和信息流程综合起来进行规划,有时候它们是合一的。

3. 信息体系结构

信息体系结构是使信息需求能与信息资源相匹配。信息体系结构通过特定的格式、类型和关系,将信息结构化。它把信息行为、信息过程、信息队伍和企业中的业务过程、组织结构以及物理场所联系起来。这里还应包括:

(1) 所有权与共享范围。由于信息是整个企业的财富,应该明确哪些是只限于某些部门和人员使用,哪些可以在更大范围内使用,哪些是由指定人员修改与更新等。

(2) 信息管理。有关的信息如何分工管理,哪些由信息技术部门集中管,哪些由部门或用户分散管。

(3) 安全与方便的考虑。为了安全,有时就要影响到使用的方便性。应该按照企业与信息的具体情况,确定适当的原则。

4. 设备体系结构

设备体系结构要考虑:

(1) 设备布局。企业中使用的是多终端集中系统还是分布式系统,它的物理布局涉及成本、使用方便性和安全诸多方面,因此对主机、外围设备等的安放要有考虑。

(2) 主机或工作站与服务器的选择。这里包括处理能力、通信要求和输出形式等的

设想。

(3) 操作系统与数据库的选择。这关系到软件的大局，要预先加以规划。

(4) 统一的标准规范（代码、数据和软件）。

(5) 应用的分布情况。网络节点的专用与通用，公用还是专用网络。

(6) 网络层次关系。包括服务范围，随处使用的可能程度。

(7) 通信线路的选择。包括频带宽度，通信协议。

2.4.3 信息系统开发管理

信息系统的开发工作是一项大型的复杂任务。除了一些小型专用系统而外，一般综合性的信息系统的开发工作技术要求高，工作量大，投资巨大而开发时间长。由于系统涉及一个组织业务的方方面面，需要的业务和技术知识面很广，涉及的部门和人员也多，因此在信息系统开始应用的阶段，许多开发工作或是以失败告终，或是距离原来要求甚远。直到后来总结了经验和教训，逐步形成了一些开发的规律，才使开发工作逐步走上正轨，信息系统的开发逐步形成了一门专业学问，在高等学校也开设了专门课程。本节只是介绍一些基本内容和步骤。

1. 系统开发的组织

信息系统的开发要有组织地进行，一般要在组织内部成立领导小组和工作组织。领导小组要有组织的最高领导参加，所谓“第一把手原则”就是必须由最高领导来发动和组织，协调各方面的工作。工作组织中要包括规划人员、业务人员、信息技术人员等。过去有一种误解，以为信息系统的开发只是信息技术人员的事。由于信息系统的工作涉及组织的各种业务，没有业务人员参加是无法完成的。

2. 系统开发步骤

信息系统的开发最好是按照项目的管理的思想与方法进行。项目的进行一般分为下面五个阶段。

1) 系统规划阶段

在项目立项之后，要进行可行性研究，结合企业总体战略确定信息系统的发展战略，定义系统的目标和功能，写出可行性研究报告，报告审议通过后编写任务书。

2) 系统分析阶段

根据任务书对现行系统进行调查，对现行系统的业务流程进行分析，找出不足之处，确定新系统的目标和功能要求，编写系统说明书。这阶段要明确“做什么”的问题。

3) 系统设计阶段

系统设计阶段要明确“怎么做”的问题，根据系统说明书规定的功能要求，具体设

计技术方案,编写系统设计说明书。

4) 系统实施阶段

系统实施阶段的任务是实现所设计的系统,包括设备的购置、安装、调试,软件编程和调试,建立数据库和网络,以及整个系统的联调,等等。

5) 系统运行维护阶段

系统测试完成投入运行后,还需要不断进行检测和维护,发现问题和缺陷及时改正。再就是对使用者进行培训。

3. 系统开发的方式

信息系统的开发,可以使用下列五种开发方式。

1) 自主开发

如果企业拥有较强的信息技术人才和队伍,可以选择由用户自己组织开发的方式。这是因为企业内部人员对自己的工作流程比较熟悉,能够准确掌握真正的需求,可以节省时间和费用。当然这并不排除请外面的专家来担任顾问。

2) 委托开发

委托开发是聘请专业开发组织来承担开发项目。这样可以在企业信息技术力量不足的情况下建成高质量的系统。但是费用会比自主开发要高。由于是外部的团队负责开发工作,对企业内部的实际情况不熟悉,因而需要企业内部的人员参加调研、分析和论证,而且需要不断地进行沟通,有时候双方还会产生矛盾,需要磋商解决。

3) 外包

企业不但把信息系统的建设委托给专业机构,同时要把日后的经常性管理维护也外包给负责开发的专业机构,这种整体性的外包,可以节约企业的资源。例如,所有的计算机硬/软件都是租用的,将来由承包方负责更新升级。

4) 联合开发

联合开发是在企业拥有一定的信息技术人员,但还无力自主开发时,聘请专业开发机构或专业人员联合进行开发工作的一种形式。这样做的好处是锻炼了自己的信息技术人员,有利于日后的运行维护,同时也节省了部分资金。只是在双方合作过程中难免会产生矛盾,需要高层领导协调处理。

5) 购买成套的系统(软件)

现在许多信息公司出售企业信息系统的所谓整体解决方案和成套设备。购买这样的成套设备的好处是可以在较短的时间内建成系统,节约时间和费用。但是这些成套系统是按照某种业务模式构建的,不一定适合本企业的实际情况,特别是一些国外的产品,在国内使用常有“水土不服”的问题,勉强使用,会有“削足适履”的弊病。好在近来

出现了许多本土产品，可供选择。而且当前的趋势是“软件作为服务（SaaS）”，供应商以外包的方式全面提供信息服务，系统也有一定的灵活性，所以这种方式会逐渐推广。

4. 系统的开发方法

常用的系统开发方法有下列三种，这里只简单介绍各自的特点，详细内容已有专门课程介绍，这里不再重复。

1) 生命周期法

生命周期法是按照信息系统的生命周期划分任务，也就是按照前面讲过的按照系统规划、系统分析、系统设计、系统实施、系统运行维护各阶段依次进行。这种按部就班方法的好处是规范有序，但由于系统开发时间过长，许多方面会发生变化而使开发工作进行困难。

2) 原型法

原型法的基本思路是开发人员凭借自己对用户需求的初步理解，先构造出一个实在的系统原型，演示给用户看，听取用户意见，反复修改直到用户满意为止。这种方法是一种渐进式开发方式，可以在双方互动过程中逐步对系统功能要求取得共识，是符合人的认识规律的，可以降低开发风险。

3) 面向对象法

前面的方法都是“面向过程”的，一旦用户需求发生变化时改动的内容太多。而面向对象的方法则是尽可能模仿人的习惯思维方式，因而接近人的解决问题方法和过程。这种方法在软件的专业课程或教材中有详细的介绍，我们这里限于课程要求不做介绍。

2.4.4 信息系统运行管理与安全管理

信息系统一旦建成，就要投入运行，向信息使用者提供信息服务。对系统运行的管理包括日常运行管理、系统运行情况记录和故障诊断、系统运行绩效的评估、系统的更新和升级、系统的安全管理五个方面。

1. 日常运行管理

日常运行管理包括：

- (1) 硬件运行与维护；
- (2) 软件运行与维护；
- (3) 信息的采集和录入；
- (4) 信息的处理与服务的提供；
- (5) 系统的安全管理。

2. 系统运行情况记录和故障诊断

系统运行情况记录和故障诊断包括：日常运行记录、系统运行情况监测、系统故障诊断和排除等。

3. 系统运行绩效的评估

系统运行绩效的评估包括：项目评价、经济评价、财务评价、行为评价和综合评价。

4. 系统的更新和升级

系统的更新和升级包括：定期更新、不定期更新和升级（由于故障或新技术的应用）。

5. 系统的安全管理

系统的安全管理包括：信息安全、计算机安全、网络安全和通信安全等。

信息系统的安全不仅影响到系统本身，而且会影响到社会经济的许多方面。因此需要特别关注。根据《中华人民共和国计算机信息系统安全保护条例》，信息系统的安全是指“保障计算机及其相关的和配套的设备、设施（含网络）的安全以及运行环境的安全，保障信息安全，保障计算机功能的正常发挥，以维护计算机信息系统的安全运行”。

为了保障安全，可以采取下面一些措施，如信息加密、数据完整性保护、身份鉴别、数字签名、访问控制、数据备份和灾难恢复、反病毒技术、网络控制技术、公证机制，等等。

2.5 信息组织与人员的管理

信息管理中还有组织和人员的问题。

2.5.1 信息管理的组织结构及其管理

对一个组织（如一个企业）来说，信息管理的组织结构模式可以有多种，其中主要的有集中型结构、分散型结构和集中分散相结合结构三种。

1. 集中型结构模式

集中型结构模式是把各种信息的采集、加工、检索和传递都集中到组织的信息中心，实行统一管理。各部门所需要的信息统一由信息中心提供和发布，信息流是辐射型传播。这比较适合中小型企业，因为可以节约人力和设备。

2. 分散型结构模式

分散型结构模式不设信息中心，它是在各部门设置专管信息的人员或者小组。组织内部定期或不定期地组织信息交流，信息流是横向传播。这种模式使得部门内的信息沟通比较迅速准确。而部门间信息流畅通与否决定于部门间的联系与合作程度。

3. 集中分散相结合的结构模式

集中分散相结合的结构模式一方面设置整个组织的信息中心；另一方面又使各部门独立管理一部分自己的信息，并建立部门间的联系。这样的信息既有纵向流动，又有横

向流动，使得信息流动比较畅通，是一种比较理想的模式，当前为许多组织所采用。

在一个比较大的企业（或者院/所）里，会有两或三层信息管理的组织机构。高层的领导机构是企业的信息化委员会或信息化领导小组。本书在前面说过，为了开发信息系统，需要在高层设立领导小组。这种小组有的在系统开发完成后就解散了。其实为长远之计，不如成立信息化领导小组或信息化委员会这样的常设机构，既管建设，又管运行。中国目前还处在工业化的中期，信息化的任务还任重道远，保持这样的组织非常必要，何况这个委员会还要管生产自动化一类的任务。这个高层领导机构应该由企业的最高层的总管（至少要管技术的总管）来领导。此外在这个层次还要设置后面就要讲到的信息总管（CIO）岗位。

高层的工作机构就是信息中心或者叫做信息管理中心，它除了组织、分析、处理、存储和发布信息外，还要协调各部门的信息工作，以及对外联系。下属的部门有系统运行部、信息服务部、技术支持部和培训部等。

中层的信息管理组织一般分布在各部门之内，专门负责该部门各种专业信息的采集、处理和发布，并负责和邻近部门的沟通。

基层的信息管理组织设在基层单位（如车间、科室），负责运作过程中的信息采集和传递，一般不处理或仅作简单处理（如汇总）。

由于信息管理的出现较晚，所以人们对信息管理机构总以为是个辅助部门，不像对财务、生产等部门那样重视。其实信息管理部門的作用有很多是潜在的，其影响是深远的，所以必须像一些主要部门那样受到关注。

2.5.2 信息人员管理

组织内部的信息管理人员包括领导人员和一般工作人员。从信息管理的发展来看，它从最早的对信息装备进行管理，发展到对信息技术进行管理，再发展到对信息资源进行管理，经历了三个阶段，而信息工作者的专业结构也随着发生变化。最早信息人员是硬件技术人员，接着就是系统（包括硬件和软件）技术人员，后来就有对信息本身进行管理的人员。

组织内常常设有信息总管（CIO）这样的职位，作为信息管理的领导人员。信息总管是全面负责组织中的信息技术和信息管理的，直接对总经理（CEO）负责。

信息总管的职责主要包括：

- （1）制定企业（或其他组织）的信息化战略。
- （2）开发各类信息化应用。
- （3）管理信息资源。
- （4）创建和管理维护企业的信息基础设施。

(5) 组织企业员工的信息化培训。

(6) 与外部有关企业建立信息方面的协作关系和进行协调。

在隶属于信息部门的一般工作人员中,中国劳动和社会保障部已经制定了有关企业信息管理师这样一种专门职业标准,把企业信息管理师定义为:从事企业信息化建设,承担信息技术应用和信息系统开发、维护、管理以及信息资源开发管理工作的复合型人才。这一职业分为高级企业信息管理师、企业信息管理师、助理企业信息管理师三级。标准要求他们掌握的基础知识包括信息技术、企业管理和法律基本知识,在专业方面的知识包括信息化管理、信息系统开发、信息网络构建、信息系统维护、信息系统运作和信息资源开发利用等。

除了隶属于信息管理部门的人员而外,还需要从业务部门中产生业务支援人才,它们应具有经营管理与生产运作业务的丰富知识与全面技能,同时又具有满足需要的信息管理与信息技术的知识与技能,能为信息部门提供有关业务方面的支援。

2.6 信息服务

信息发挥作用可以通过信息服务来实现。

2.6.1 信息服务的含义

前面曾经讲到,信息从生成、采集、处理和传递到提供使用,经历了不同的阶段。信息的提供实际上就是为信息使用者提供信息服务。其实人们在日常生活中,总是自觉不自觉地接受信息服务。

关于信息服务,目前有广义和狭义两种理解。广义的理解认为:信息服务是指所有那些以产品或劳务形式向用户提供和传播信息的各种信息活动。这里所谓信息活动,包括信息内容的产生与开发、信息的发布与分配、信息的流通与传播,以及信息提供服务和信息技术方面的服务。

狭义的理解则认为:信息服务是专职的信息服务机构针对用户的需要,及时把开发加工后的信息,以用户方便接受的形式准确传递给特定用户的活动。

在大多数情况下,信息服务者本身不生成信息,而是通过选择、加工、提供信息内容来满足用户的信息需要,成为联系信息内容和信息用户的桥梁。他们帮助用户克服信息交流障碍,解决信息生成的广泛性和信息利用的特定性之间的矛盾,使信息资源的充分开发和有效利用得到有机的统一。

对信息服务的要求包括信息的提供要有针对性,信息的提供要有适时性,信息的提供要使用户感到方便,要讲求经济与社会效益四个方面。

1. 信息的提供要有针对性

任何信息都是在特定的时间、特定的场合下对特定的需要才产生效用,因此信息服

务必须紧密围绕特定用户的具体信息需要展开，注意信息提供的针对性。所谓针对性问题，也就是服务内容和 service 对象的匹配问题。因此信息提供者要认真研究用户的需求特点和需求的变化，以及利用信息的习惯，选择适当的信息渠道、信息载体和信息形式，以满足用户需要。

2. 信息的提供要有适时性

信息具有时效性，且时间对于信息的价值具有决定性影响。因此最好把握恰当的提供时机，在最需要的时间把最需要的信息传送过去。传递过晚会使信息失去价值，而过早则不容易引起重视。

3. 信息的提供要使用户感到方便

根据对用户的调查研究结果，用户选择和利用信息时，信息的可获得性和易用性是最重要的因素。信息服务要为用户的信息活动提供最大的便利条件，如使用信息技术克服时间和空间的障碍，简化信息服务的手续，提高信息服务的效率，开发用户友好的界面，开展用户培训，等等。

4. 要讲求经济与社会效益

对用户来说，支付了一定的费用，花费了一定的时间和精力，应该保证它们获得一定的经济与社会效益。对提供者来说，应该尽量提高效益，降低费用。

2.6.2 信息服务的方式

目前信息服务的方式有文献服务、报道服务、信息检索服务、咨询服务和网络信息服务五种。

1. 文献服务

文献服务是历史最悠久的传统服务方式，如图书馆、资料室、档案馆等就是以这种方式为主的信息服务机构。具体的服务方式都是人们非常熟悉的：阅览服务、外借服务和复制服务等。

2. 报道服务

信息服务机构把采集到的信息内容经过加工整理与分析综合，形成报道以便于用户使用。具体的形式有：口头报道，如信息发布会；直观报道，如展览会；文字报道，如信息出版物。

3. 信息检索服务

信息检索服务是根据用户要求，由专门人员辅助或代替用户查找信息并将结果以一

定形式提供给用户的一种服务。通常有：文献检索、数据检索和事实检索。关于检索的方式方法本书在后面还会介绍。

4. 咨询服务

咨询服务是指咨询机构或者咨询人员以专门的信息、知识、技能和经验，运用科学的方法和先进的手段，对委托方提出的问题提供一种或者多种解决方案或提出建议。这包括技术咨询、工程咨询、管理咨询和政策咨询等。

咨询服务实际上是一种知识服务，常常需要进行调查研究、分析综合、判断与决策建议等，其中许多工作是信息活动。

5. 网络信息服务

网络信息服务是指在网络环境下，信息服务机构利用计算机、通信网络等现代设施，为用户提供信息采集、处理、存储、传递和应用等服务。这是一个包含内容丰富、覆盖面极为广阔的服务，特别是互联网上的服务，现在大多数人已经享受这种服务了。

2.7 信息管理中的人的因素问题

最后，再来探讨一下信息技术应用中的人的因素问题^[4]。

2.7.1 信息生态

在相当长的一段时期内，人们在信息技术应用的研究、开发和应用进程中，大都把注意力放在技术上，尤其是放在计算机系统上，形成了人围着计算机转的局面。这在计算机开始应用的最初阶段的确是不可避免的，对于当时的主机这样一个庞然大物，构造复杂，故障时常发生，需要人们花费很多精力去应付。但是到了20世纪80年代以后，由于计算机系统集成化、模块化的发展，硬件的可靠性大大提高，随着软件、特别是系统软件的进一步完善，对计算机系统的维护工作已大大减轻。而人-机接口的发展，促使非计算机专业人员已能自己运行微机系统，但人们的注意力仍集中在不断换代更新的技术装备上，忽视了信息系统中的主要内容——信息，忽视了信息应该怎样为人服务。

因此，研究信息系统怎样能为人所有效利用，变“人围着计算机转”为“计算机（信息工具）随着人转”，是一个带有根本性意义的问题。

美国的管理科学家Davenport提出了信息生态学的概念^[4]。他认为应该着重研究一个企业或其他组织的整体信息环境，研究人是怎样生成、分发、理解和使用信息的。

在他看来，信息生态学应该探讨下列几方面的问题：

（1）组织内部对信息的价值观与信念问题，这属于企业文化问题。

(2) 人们实际上怎样使用信息, 对信息做什么工作的问题, 这属于人的行为与工作过程问题。

(3) 人对信息的支配权力, 如集权与分权、共享等问题 (Davenport 称为信息政治)。

(4) 技术本身的问题。

关于信息环境、信息生态学的研究, 无异应该以人为中心。在一个机构中, 信息每时每刻都要受到权力、政治和经济的影响。这已经不是什么秘密了。但是很少有管理者希望去有意识地、系统地去涉及这个问题, 这也许是怕因此影响或割断了组织中已经建立起来的层次结构。在许多组织中, 信息支配权 (信息政治) 是不可讨论的, 但却存在着内隐的信息支配权结构。其中的某些权力活动的负面影响会使信息与知识的应用发生变形。

许多人把注意力放在信息技术的支配权上, 如考虑企业信息主管 (CIO) 的作用、信息技术特色的管理等, 但实际上同样重要、甚至更加重要的却是对信息本身的支配权问题。

在企业中, 谁来决定应该收集什么信息, 怎样有效地使用信息, 要比使用任何新技术都重要, 因为它能改变企业的运作。

当前有信息联邦模式、信息分封模式、信息君主模式和信息无政府模式四种信息统治或者说支配的模式, 这都是对比现实政治体制来命名的。这也体现了从集权到分权的一个连续过渡: 君主模式最集中, 联邦模式次之, 再次是分封模式, 无政府模式是极端分权的。

下面分别研究一下这几种形式。

1. 信息联邦模式

信息联邦模式的中央集权较弱而地方自治性较强。在信息方面, 则是只有少量信息元素需要集中定义和管理, 其余都可由各个局部来定义和管理。这当然需要总部和各局部之间达成协议。

国外有的公司认为, 三种信息要集中管理: 一是支持法律法规所需的信息, 二是高层领导指定下属所完成的职责的信息, 三是对多个业务领域有关部门之间的信息。

当然, 联邦模式并不是任何企业都适用的, 这要看企业的规模与经营范围。对于小型、单一业务的企业就不值得采用。

2. 信息分封模式

信息分封模式是指企业内各部门各自管理自己的信息环境, 就像各路诸侯统治自己的领地一样。这是当前遇到的最多的情况, 但当事人最不易觉察和探究。它虽不及联邦

模式在大企业中那样健康地生存，但它毕竟也已形成并大量存在。分封体制允许各路诸侯使自己的信息更贴近自己的业务需要。有时这也会引起高层领导的不满意，但各单位实际上已经是心照不宣地在这样做了。

3. 信息君主模式

如果企业中大多数的信息控制在一个人或一种岗位职能者手中，那就是信息君主模式。君主可能就是某一高层领导，也可能不是。他指定什么信息是重要的，确立关键性信息的意义，并控制对信息的解释权。这对小规模、单一业务的小企业是有利的。

4. 信息无政府模式

信息无政府模式不是实际上稳定存在的一种模式。在这种情况下，每个人只管自己，这并不是企业想这样做，常常是当企业想实现集权遭致失败，或者根本没有高层领导了解分享信息的重要性时发生的情况。

微型计算机的广泛应用也为这种无政府状态提供了可能性。一些小单位或者个人突然发现，不需要很多的投资，就能建立自己的数据库，在任何时候按自己的需要裁剪信息。在几乎所有的无政府模式下的企业中，都有大量的知识工作者，如科学家、咨询专家、程序设计人员和系统开发人员等。因为他们已经把信息看作是对自己工作有影响的事情，而一旦要采用一种更有效的模式，他们能够较快地进入。

但是这种信息无政府模式的缺点也是很明显的。每个人有自己的数据库，各人的数据库中的数据值，因所在部门不同而有所差异，因而会造成报表的不准确性。

一个企业如果总是处在这种状态下是难以生存的，必须向另一种适当的模式转化。也有的企业，上述四种模式都不适合，如企业中的信息需求确定了信息的支配权力。这可以认为是一种面向市场的模式，由市场来逐步形成它的内容。当然这需要有能力确定实际需求。它的缺点是对信息的所有权不明晰。

有人迷信技术可以解决一切问题，以为一个新的设备或工具的出现就能解决经营管理中的问题（现在有些技术产品公司就是这样宣传的），实际上这是完全不可能的。

上面所说的信息支配权是指战略上的。有时候人们也在支配权力上采用一些战术行动。例如，信息交易和信息泄露等，有时是公开的，有时是秘密的，这些活动有的会削弱信息支配权，有的会强化信息支配权，不可不加以注意。

尽管 Davenport 讲的是西方国家的情况，但在中国当前市场经济形势下，企业中的信息支配权也或多或少地与上面讲的情况相似，过去计划经济体制留下的信息支配权过于集中的问题仍旧存在，所以对这一问题应该十分重视。因为任何体制的核心问题，都涉及决策权、信息支配权与利益分配权，而三者之间是相互关联的，其中信息支配权最为隐蔽。

2.7.2 信息用户利用信息的心理机制

用户利用信息的心理活动，包括两个方面，一个方面是信息利用中的认识活动，另一个方面是信息利用中的意向活动。现在分别加以研究。

1. 信息利用中的认识活动

人在信息利用中的认识活动大致可以分为认知阶段、理解阶段、思维和评价阶段、应用和创新阶段四个阶段。

1) 认知阶段

人的认识活动是从感知开始的，感知是认识客观事物的最初源泉，是吸收信息的第一步。感知阶段又可分成感觉和知觉两个方面。感觉是对客观事物个别属性的直接反映。由于信息经常以文字、符号、图形、话语形式或者实物形式表现出来，因此感觉活动首先是用眼睛看或用耳朵听，其次由视神经或听神经对广播或声波传来的信号进行辨认分析，传给大脑皮层。

知觉是对直接作用于感官的客观事物整体的综合反映，将感觉到的信息组成有意义的对象，再与头脑中已有的信息、知识、经验相比较，形成一般性的认识。

感知信息一般有两个层次，一个是发现，寻找必要的注视点，这与人的文字水平、知识水平、信息能力有很大的关系；另一个是辨别，将有意义的信号与无意义的信号、关键的信息与次要的信息分辨、区别开来，这是知觉的选择过程。

2) 理解阶段

在利用信息的过程中，理解就是运用已知的知识去掌握新的知识。理解的过程是从感性认识上升到理性认识的过程，是对信息内容进行分析和吸收，并将信息内化为个人的观念和知识的过程。理解的目的在于解释信息表述的“是什么”和“为什么”，以及信息与主体、环境之间的关系。这样就形成了新的知识。

3) 思维和评价阶段

思维是对信息中的知识进一步分析、比较、判断或推理，从而对信息中的知识有更深入的认识与概括。经过人脑的思维，所反映的已经不是事物的个别特征，而是同一类事物的本质特征和事物之间规律性的联系。

目前，人们认识到的有意识的思维有三类：抽象（逻辑）思维、形象（直感）思维和灵感（顿悟）思维。

抽象（逻辑）思维是人类智力与神经系统发展到高级阶段才产生的，它是人们透过现象认识事物本质的重要手段，经过多少年的努力，出现了形式逻辑、辩证逻辑、数理逻辑等一系列学科，推动了自然科学与社会科学领域中显性知识的生成和学科的发展。

我们日常所谈到的像分析与综合、抽象与概括、归纳与演绎、比较与归类都是抽象思维的方式。

形象（直感）思维与抽象思维一样，也是人类理性认识的一种方式。形象思维从整个人类来说它们都是在感性认识的基础上开始的，但在继续前进时刻走上了不同的道路。抽象思维是指对事物的间接的、概括的认识，用抽象材料（概念理论、数字等）进行思考；而形象思维则主要用典型化的方式进行概括，并用形象材料来思考。两者虽然使用的材料不同，思考方法不同，但都可以认识事物的本质。有人认为形象思维不过是一种直观的认识，只能是一种感性认识。其实不然，因为在形象思维过程中作为思维材料的形象，已不是原始形象，而是经过加工改造了的形象。

灵感（顿悟）思维涉及人的灵感似乎是很神秘的东西，但近年来心理学家认为它与人的潜意识有关。有一种说明和解释灵感思维的假说，认为灵感思维是人的显意识与潜意识的相互作用。所谓潜意识，乃是未显示的意识，是人脑不可缺少的潜在的反映形式，是主体对客体的信息的前控制及内部体验的统一。

4) 应用和创新阶段

应用和创新阶段要把从信息中获得的知识进一步应用到实际生活中去，解决实践中的问题，或者进行创新。创新是根据一定的目的和任务，运用已有的信息和知识，进行能动性的知识活动，借助于想象、灵感和思维互动等，以渐进或突变形式产生新概念、新理论或者新设想和新的发明创造。

2. 信息利用中的意向活动

信息利用中的认识活动是人类的一种普遍的心理活动，而信息利用中的意向活动却是带有较多个人心理色彩的心理活动，受人的先天特性和社会条件影响与制约。

信息利用中的意向活动有信息动机、个人兴趣、个人情绪和个人意志四种。

1) 信息动机

信息利用的动机是一种内在动力，反映了个人愿望和个人需求。它的作用表现在信息利用的方向指引、信息行为的强化上。了解用户的信息动机，可以掌握用户在信息利用中的心理状态和倾向。

2) 个人兴趣

个人兴趣反映的是用户的爱好程度，决定于客观上的工具和运用机制的方便程度，以及主观上的能力与习惯。对用户兴趣的掌握有利于信息活动的顺畅进行。

3) 个人情绪

个人情绪是用户从事信息活动时心理的外在表现。情绪能激发用户的积极性，促进注意、记忆、联想和创造性思维，但也可能对信息活动造成障碍。在信息活动前、活动中与活动后的情绪会有变化，如果掌握了并引导了这种变化，会增强信息的作用效果。

4) 个人意志

个人意志是用户自觉确定目标、支配自己去达到目的的心理过程。个人意志可以使得用户克服困难、排除干扰、专心从事信息活动,表现出自觉性、自我可执行和持久性。

上面说到的这些心理活动已经不仅限于狭义的信息活动,而且涉及后面将要讲到的知识的获取与生成及创造力的发挥。对于这些心理活动的关注正是体现了以人为本的管理思想。

2.7.3 信息素养

信息素养(Information Literacy)是指人在信息社会中生活所应具备的信息意识、信息能力和信息道德修养。

1. 信息意识

信息意识包括用户对信息和信息环境的感觉、知觉、情感和意志等,感觉、知觉和思维是信息意识的源泉和基础。信息意识表现在:

(1) 对信息具有特殊敏锐的感受力,能够敏锐地捕捉信息,从别人忽视的现象中发现有价值的信息。

(2) 对信息有持久的注意力,对信息的关注不受时间和空间的限制,随时随地习惯于用信息眼光观察周围的一切事物。

(3) 对信息价值的判断力和洞察力,能够透过信息的表面现象深入发现其本质特征,能够去粗取精、由表及里。

2. 信息能力

信息能力总的来说包括:

(1) 利用新的信息技术来获取和处理所需信息的能力。

(2) 当遇到问题需要某种信息时主动去查找和使用信息的能力。

(3) 掌握查找信息的途径和方法,具备组织、分析、鉴别、评价信息价值的能力。

在信息活动中,要具有:

(1) 信息获取能力,即能够积极主动从各种渠道获取所需信息,这是信息能力的基础。

(2) 信息加工能力,即能够根据需要筛选信息,去掉冗余信息,以各种方式对信息进行加工,对信息加以组织使其有序化,以提高信息的价值。

(3) 信息的利用能力,即将自身原有的信息与新的信息集成,经过综合,转化为新的信息或知识。

(4) 信息技术的应用能力,即在新技术出现时能够很快地加以利用,提高信息的获取、处理、存储和传递的能力。

3. 信息道德修养

道德是在一定的社会中调整人们相互关系以及人与社会关系的一种行为规范。信息正是这种联系的手段,因此它与道德有着不可分的关系。

对于信息生成者和信息服务者来说,在信息道德方面要做到:本着对社会正义和进步的责任感,尊重客观事实,尊重学术道德规范,不弄虚作假,抵制错误信息,遵守国家法律。要积极保护知识产权,尊重个人隐私。

对于信息使用者来说,要尊重他人的创作权和信息所有权与隐私权。不进行信息偷窃,不利用信息作伪证,不在未经允许的情况下使用他人的信息资源。在利用信息时,要考虑其社会后果。要积极以新的信息和知识为社会谋福祉。



思考与讨论题

- (1) 信息处理和信息管理有什么差别?又有什么联系?
- (2) 信息与信号有什么差别?两者又有什么联系?
- (3) 你能举出一些不用信息技术工具的信息系统吗?
- (4) 你对现在已经有的信息技术系统的发展前景有什么预见?
- (5) 你觉得个人信息也需要管理吗?
- (6) 试列举一个科研机构中信息资源的类型与存在的形式。
- (7) 通过什么手段来减少和避免信息犯罪?



参考文献

- [1] 杨善林等. 信息管理学[M]. 北京: 高等教育出版社, 2003.
- [2] O'Brien J A. Introduction to Information System[M]. New York: McGraw-Hill Inc., 1998.
- [3] 黄梯云, 李一军. 管理信息系统(第4版)[M]. 北京: 高等教育出版社, 2004.
- [4] 王众托. 企业信息化与管理变革[M]. 北京: 中国人民大学出版社, 2001.
- [5] 谢阳群等. 微观信息管理[M]. 合肥: 安徽大学出版社, 2007.
- [6] 杨志芳. 信息管理基础[M]. 西安: 西安交通大学出版社, 2008.
- [7] 钟义信. 信息科学与技术导论[M]. 北京: 北京邮电大学出版社, 2007.
- [8] 刘红军. 信息管理概论[M]. 北京: 科学出版社, 2008.
- [9] 柯平, 高洁. 信息管理概论[M]. 北京: 科学出版社, 2007.
- [10] 杜栋. 信息管理学教程(第三版)[M]. 北京: 清华大学出版社, 2007.

第 3 章

知识管理概述

内容提要

本章是对知识管理的概括介绍。首先介绍有关管理的含义与任务，其次介绍系统工程思想在信息管理中的应用。然后从知识资源、知识组织与人员、知识系统的基础设施、知识经营与知识服务、知识的社会文化环境各方面阐明知识管理的特点。

本章重点

- 知识管理的定义与特征
- 知识系统工程
- 知识系统的体系结构
- 知识管理的非技术因素（组织、人员、社会文化环境）



3.1 知识管理的概念

我们先从知识管理的含义与特点讲起。

3.1.1 知识管理的含义

近些年来，知识的应用日益广泛。人们一方面在采集、处理、使用和创新知识，同时也在自觉不自觉地在组织和管理知识。对那些有固定形式的知识，如技术文化资料、图纸、专利等一般还能自觉加以管理（即通常所谓的技术管理或档案管理），而对一些无形的，特别是隐性的知识，则常常未予重视，因此在人员退休或者离开企业后，有些极宝贵的隐性知识无形中就散失了。即使是有形知识，由于散在各处，缺少有效的索引，常常出现重复引进或开发，或者出现传播过于迟缓的情况。现行的企业制度、风气甚至有时会损害或妨碍知识的获取、传播和应用。因此，把知识管理作为一个重要的任务提出来，加以认真地理解和研究是有现实意义的。

现在企业界已经越来越重视知识的管理。**Baldrige** 奖励基金会曾经针对今天的社会经济发展，向 300 位企业总管征求有关什么是当前企业发展的重要趋势意见，有 88% 的人认为知识管理是重要趋势，仅次于国际化（占 94%）而居第二位。

各级政府在社会经济形势急剧变化的复杂条件下，也感到及时正确的决策需要掌握更多的知识。人们常把决策失误归咎于缺少信息，实际上缺乏的是能够洞察全局和预见未来的知识。至于科技与教育界对知识的依赖就更加迫切了。因此对于知识这种资源和生产要素的管理，也就提到日程上来了。

近年来，人们在研究知识管理的过程中，对知识管理做出过下列各种定义^[1,2]：

“知识管理是将所有的专业知识，不论是在纸上，在数据库里还是在人的头脑中，加以掌握并分配到能够产生最大效益的地方去。”

“知识管理是获取恰当的知识在恰当的时候交给恰当的人，使他们能做出最好的决策。”

“知识管理涉及发现和分析已有的和需要的知识，并规划和控制开发知识资产的行动，以达到组织的目标。”

“知识管理就是鼓励创新与知识共享。”

“知识管理是系统地寻求、处理、理解和使用知识以创造价值。”

“知识管理乃是利用组织的无形资产创造价值的技术。”

“知识管理乃是以信息为基础的活动，通过组织性学习创造显性与隐性知识。”

“知识管理乃是运用集体的智慧提高应变和创新的能力。”

“知识管理可以帮助人们对拥有的知识进行反思,调整和发展企业内部结构,提高人们进行知识交流的技巧,增加获得知识的来源,促进知识交流。”

上面这些定义从不同的角度说明了知识管理的任务、内容和特点。我们不必去追求一个统一和全面的定义而可以从知识管理的特点和具体任务来进行理解。

3.1.2 知识管理的特点

工农业生产的管理,是对实物生产活动的管理,而知识管理则是对知识活动的管理。知识活动是对一种特殊的、抽象的要素——知识加以处理或者加以创新,得到的知识又使用到实物的生产或服务中去。如果说在实物生产中,生产过程和管理过程形成了两个不同层次的话,知识处理(如研究开发、设计等)与知识管理也形成了两个不同的层次。

在实物生产活动中,人们要对原料、加工设备、加工过程、产品等这些有形的对象加以管理,而在知识的应用和创新活动中,人们则要对知识与处理知识的人和设备进行管理。由于知识是一种无形资产,所以它的管理远比有形物资管理来得复杂。特别是隐性知识存在于人的头脑之中,人员流动又会造成知识的散失。

考虑到知识管理的特殊性,在组织和管理知识时,应该充分注意到下面一些特点:

(1) 知识产生并存在于人的头脑之中,知识运作必然是以人为主导的,人的价值观、信念、情绪和意志都将会产生影响。

(2) 技术工具能增强知识活动,但工具的正确与有效应用还决定于人。

(3) 必须同时重视显性知识与隐性知识,两者不是对立的,而是互补的,当前尤其应该重视隐性知识的获取与转化。

(4) 知识共享是创新的必要条件,应该创造条件鼓励和推进知识共享。

(5) 知识管理必须受到高层领导的重视,并列入组织的经营管理总体框架之中。

(6) 知识管理是在一定的文化背景基础上进行的,因此民族文化、企业文化的各种因素常常成为决定性因素。

目前对知识管理有宽窄两种理解,一种更加宽泛的理解是把知识管理当作一种新的管理理念。这种理解认为,在知识经济时代,一切管理都应该以知识为基础。另一种较窄的理解则认为知识管理仅仅是对企业的知识生成与应用加以管理,就像财务管理一样,只是企业管理的一个领域。前者可以说是“用知识进行管理”,后者可以说是“对知识进行管理”。在目前管理界所理解的知识管理还是指后一种而言。

创新要依靠组织(企业、院所)的知识,而怎样使创新集体中每个人的知识都能与其他人沟通共享,形成组织知识,是一个非常复杂的组织行为和社会过程。传统的管理思想和模式在这里是不适用的,需要另辟蹊径。

当前由于知识管理刚刚兴起,所以还有许多人对它有片面的或者不恰当的看法。

有些人认为知识管理只是对现存的知识进行管理,以利于多次使用,而没有意识到,更重要的是对怎样生成新知识进行管理。

有些人认为知识管理仅仅是利用信息工具对可编码的知识进行管理,没有意识到,除了可编码的知识外,隐性知识在知识管理中是十分重要而且又是无法用信息工具直接管理的。

也有人认为知识(特别是隐性知识)存在于人的头脑之中,别人无法觉察,是无法加以管理的。

凡此种种,都使人产生误解而对知识管理的重要性认识不足。

由于知识管理的高度复杂性,兼有技术与人文两种属性,而且两种属性是交互作用的,在知识管理的实践中,不仅有技术过程,而且有组织过程,特别是人际沟通过程。因此知识管理具有跨学科特点,涉及的学科有:管理学、经济学、认知科学、心理学、社会学、信息科学与技术、系统科学与系统工程、法学和哲学,等等。

由于知识管理涉及多种学科,所以在知识管理的发展过程中,一些名词、概念和定义等还要借用所涉及的领域中已有的结果,但不同领域的定义、概念会有所差异,这也会给知识管理的研究带来困难,因而特别需要横跨几个学科进行开拓,在原有学科的边缘建立新的领域。

3.2 知识资源、知识资产和知识资本

这一节将依次介绍知识作为资源、作为资产和作为资本的一些概念和特征。

3.2.1 知识资源

知识资源是指人类智力劳动所能发现和创造的成果,它可以被人们用在物质产品生产或精神产品生产之中,创造新的价值。

我们通常说到的资源,在不同的采掘、加工阶段有不同的理解。从资源产业(如煤矿、铁矿)来说,资源是指处在尚未被人们准备加工、利用的状态或阶段的资源,如地下的煤层。而一旦开采出来,煤炭就是产品了。但是从发电厂来看,煤炭是资源,发出的电是产品。信息和知识资源也是这样。对于采集原始信息与知识的人或者组织来说,那些尚未被发现和采集的应该算是资源,而一旦采集到手,就变成了有一定意义的信息内容或知识内容。但是对信息或知识加工的人或组织来说,采集到的信息或知识又是待加工的资源。因此在实际生活中对资源的理解就显得非常宽泛。我们在后面将要说到的信息或知识资源,就带有这样的宽泛特点。但是一旦人们自觉认识到知识并获取、掌握了,就变成了个人或组织的知识资产了。将知识资产投到经济活动中去,以获得利益,

就变成了知识资本了。

知识资源有狭义和广义两种理解。狭义的知识资源仅指知识本身，广义的知识资源是指除知识本身而外，还包括知识产品（如书刊和软件等）、知识工具（如知识管理系统）以及其他一些无形资源（如声誉和形象）。

前面曾经讲到，对于知识的研究，可以从它的存量和流量两方面着眼，在研究知识资源的过程中，也可以从这两方面加以考察。

知识存量是指某一个组织在某一个时刻对知识资源的拥有量。图书馆中的藏书就是这一类。知识存量是一个静态的概念，但会随着时间逐渐变化，如人类当前所具有的知识，一般说来是与日俱增的，但也要推陈出新，也有一些过时的知识会逐渐消逝。

在知识经济活动中，由于人所创造的新知识主要是应用性的知识，所以知识存量主要是指应用知识的资源存量。

按照联合国经济社会委员会的看法，这类存量包含设备、人力、信息和组织四种要素。

1. 设备要素

设备要素是指生产的工具与设备，是技术的实体。知识已经嵌入到这些工具与设备中了。其中嵌入的不仅有显性知识，而且有隐性知识。人们虽然很容易认识这些实物，但并不太容易认识其中的知识含量。

2. 人力要素

人力要素是指人的生产技能和经验，是技术的人为要素。这里特指依附在人身上的隐性知识，因为它是无法用其他要素替代的。

3. 信息要素

信息要素是指有关的信息与档案资料，是技术的信息形式，其中的知识主要是显性（可编码的）知识。

4. 组织要素

组织要素是指生产的组织、计划与管理，是技术的组织形式。这里包括组织机构、管理制度、决策程序以及协调方法等。知识已经嵌入到上述组织或过程中了。其中既有显性知识，又有隐性知识。

在这四种要素中，人力要素是知识应用与创新的主体；设备要素是物质保证；信息要素代表了知识的积累；组织要素则是把人力、设备、信息组织起来，协调它们的工作，支持它们的活动。四种要素相互关联，相互支持，成为知识资源的存量整体。

知识流量应该被理解为在某一时段内流入和流出系统的知识资源（包括可编码的知识和隐性知识）的数量。知识流量是一个动态的概念，一般所说的是流入/流出整个组织

的知识数量。但是在组织内部，各部门之间也有流量。通常考虑的是组织整体从外部的输入/输出流量，它包括：

(1) 组织与公共知识源（如互联网和数字图书馆等）之间的知识流。

(2) 组织与其他组织之间的知识流。例如，企业与企业之间、企业与科研院所之间的知识交换。

(3) 购买或出售专利或技术许可。

(4) 人员的流动和人员的培训。

知识流动的类型多而涉及范围广，流动速度又很快。最初人们多采用另外两个代用指标：一个是物化形式的知识扩散，即新技术知识采用到装备和过程中的速度；另一个是非物化形式的知识扩散，也就是知识与专长以专利、许可证的方式进行的技术传播。但这远远不足以真实反映知识的流动情况。

物化形式的知识流量，在具体生产领域可以用特定技术在各部门的扩散与使用情况来度量。非物化形式的知识流量目前只能通过文献的引文分析来进行测度。因为无论是在学术期刊上发表或者是专利申请，知识和概念的使用者都要引述其出处，这就有可能在特定领域中使相关的思想和概念相互联系起来。专利是知识创新活动的载体，通过专利的应用追踪可以测度知识在各领域之间的流通情况。

3.2.2 知识资产

资产是一种能为组织创造价值、带来经济效益的资源。由于知识能为组织带来经济收益，所以知识也是组织的一种资产。本书前面曾经提到，在知识经济时代，由人的知识和能力构成的知识资产是经济发展的主要动力。

由于知识有显性知识和隐性知识两大类，所以知识资产也有显性知识资产和隐性知识资产之分。个人所有的显性知识资产和隐性知识资产可以统称为人才资产，一个组织（如一个企业、一所院校或科研机构）所有的显性知识资产可以称为知识结构资产。组织所有的隐性知识资产中一部分由个人的隐性知识资产构成，会随着人员而流动；另一部分已经融入组织的工作习惯之中，也是一种结构资产。

知识成为资产，首先是从直接用于生产的技术知识开始的。在社会经济发展水平相当低的历史时期，人类劳动主要是以个体小生产的方式进行的，社会分工和协作是很少的。当时绝大多数的生产知识是在生产实践中积累而形成的，主要处于隐性知识阶段，只能以师傅带徒弟的方式传授，无法加以提炼形成系统的知识加以传播。到了产业革命前后，由于技术和机械的广泛应用，生产过程日益复杂，社会劳动分工不断细化，知识要素逐步从其他生产要素中剥离出来，以科学技术研究开发为核心的知识生产过程逐渐与生产过程分离，知识生产成为一种专门的职业。知识生产者的产品既不是消费品，也

不是直接的劳动资料，而是知识。这些知识的价值虽然不具有物质性，但却是为物质生产者所需要，用来提高经济效益或者提高社会效益的。由于知识生产的专业化和知识的可以重用，因而使知识逐渐脱离实物资产而形成一种新的资产。

知识资产的聚积和壮大，加速了工业经济向知识经济的转变，对知识经济起到了催生作用。高科技产业的产生与蓬勃发展，是靠知识资产的运用，而高科技对传统产业的更新改造也主要靠知识资产来进行两者的集成。人们可以借助于智力与信息，提高生产和创新的能力，并可使同样的自然资源创造出更多的物质财富。人们在不断满足物质需要的同时，也努力创造出精神食粮，满足精神升华需要，巨大的精神财富正是智力劳动的产物。

随着知识经济的发展，企业的效益越来越依赖于知识与创新，因此企业的管理重心，也就逐渐由生产转向创新，企业的投资方向也由固定资产转向知识资产，不但在研究开发方面增大投入，而且在专利、技术引进以及人员培训等方面也增加了投入。中国在新型工业化的发展过程中，有些企业常常热衷于固定资产的投资，在厂房、设备等方面考虑得多而在知识资产方面想得少，技术常靠全面引进，忽视智力资源的投入与合理使用，由于不掌握核心技术和没有原创性的技术而受制于人。长此以往，会使发展速度和国力水平受到影响。当前强调自主创新，正是为了摆脱这种处境。

对于国家来说，不但要拥有长期历史积累的丰富的知识资产，而且要不断地获取和生成新的知识资产，并且利用这些宝贵的知识提高社会经济发展水平和综合国力。国家现代化的水平表现在国家对知识资产的有效运用和创新的水平。

对于个人来说，知识资产的积累和有效应用决定了生存和发展的能力和成就，因此人人都需要不断地学习，真正做到活到老学到老。

3.2.3 知识资本

知识资本可以这样定义：知识资本是将个人知识资产或组织知识资产以分配契约的方式投资到企业中去，目的是为了获得这项投资的企业经济剩余分配权。这时知识资产就成为知识资本。

知识资产由知识资源转化而来，而知识资本又是由知识资产转化而来。知识资源、知识资产、知识资本是三位一体的。

当知识资产进入经济领域时，如果把它作为一种获得企业经济剩余分配的投资，它就变成了知识资本。如果将知识资产作为商品进行交易（如技术贸易、专利贸易等），主体只能获得交易价值。知识只有转化为知识资本以后，才能获得巨大的收益。知识资产用于公益性服务，获得的是社会效益。

知识资本是一种新的资产形态，它是无形的、潜在的，不同于传统的实物资本与货

币资本，但它要依附于资本权属主体与一定的社会环境，主体消失了，知识资本也就失去意义。不像实物与货币资本，主体消失了还可以转移。知识资本又是一种虚拟资本，它不能作为一种独立的资本进行经营，必须在货币资本的协作下才能获得投资的收益，才能体现知识资本的价值。但是知识资本具有一种独特的性质，就是创新性。因而在经济活动中具有增值性。它的流动性很强，而且有相当大的灵活性，适应于不同环境，以不同的方式运行。

对知识资本所包括的内容，当前有许多不同的看法。比较得到多数人认可的是包含三种资本，即人才资本、结构资本和市场资本。

另外一种看法是把知识产权资本从结构资本中分出来，形成四种资本。由于当前人们对知识产权比较关注，所以四分法有它的优点。

下面将按四分法逐一介绍各种知识资本的含义。

1. 人才资本

人才资本在有的著作中称为人力资本，但是由于一般理解的人力资本中还包括体力劳动部分，为了专指智力部分，称为人才资本有利于区别。人才资本是由企业员工所具有的知识、技能和学习能力、创造能力、完成任务能力等智力因素所构成的。它是知识资本中最活跃的因素，是联系其他部分的纽带。它是知识密集型企业实现价值和价值增值的重要基础。我们今天不但把人的智力劳动看作是一种生产要素，而且看作是一种资本，是因为它具有无限增值的可能性。只有企业员工把自己的知识、能力发挥出来，与结构资本、市场资本相结合，才能使知识资本增值。缺少人才资本，其余两部分是无法发挥作用的。

具体说来，人才资本决定于企业中人才的如下方面：

- (1) 基础教育状况；
- (2) 专业教育状况和专业资格认定状况；
- (3) 与实际工作密切相关的知识；
- (4) 实际工作能力；
- (5) 个人性格和心理素质。

2. 结构资本

结构资本由企业的组织结构、制度规范、管理系统、信息网络、组织文化等组成。有人认为，结构资本是一个企业的“骨架”，它以企业基础结构为依托，合理配置各种实物资本、人才资本、市场资本、知识产权资本，使企业运营起来，产生效益。

结构资本具体的要素有：

- (1) 企业所有制与组织结构。它是一个企业管理思想的具体体现，它能够把散乱的

生产要素配置成一个有机的整体。

(2) 企业的管理系统。这是企业为了达到自己的目标而建立的一整套规程和工作习惯体系。

(3) 企业的信息网络。信息网络不仅包括由信息技术组成的互联网、内联网和外联网,还包括传统的沟通工具。

(4) 企业文化。文化是指社会或共有的价值观和行为的标准。企业文化乃是以企业整体的价值观为基础、对员工日常行为起规范作用的文化氛围。

3. 市场资本(关系资本)

市场资本(关系资本)乃是支持知识资本转化、实现生产目标的手段,它包括客户、品牌、商誉和营销渠道四个方面的内容。

(1) 客户:客户是使用企业生产的产品或者接受企业提供的服务的群体。客户分为潜在客户、一般客户、忠实客户这几类。潜在客户是有可能使用企业的产品或服务的个人或组织。一般客户是偶尔使用企业的产品或服务的个人或组织。企业应该使他们越来越频繁地使用本企业的产品或服务,进而产生对企业产品和服务的依赖,成为忠实客户。忠实客户则是对企业的产品或服务绝对信赖,只使用该企业的产品或服务。企业与客户的关系是建立在隐性知识之上的。

(2) 品牌:品牌是一个名字、术语、标志、符号或者它们的组合,用来识别某个或某些销售者的产品或服务,并因此区别于其他竞争者的产品或服务。对客户而言,品牌可以使他们在众多的同类产品中迅速识别某一企业产品的标志。品牌是技术知识、营销知识的凝聚,它的价值常常是无法估量的。

(3) 商誉:商誉是指企业在长期的运作过程中形成的信誉。它与企业的客户资本、品牌资本互为因果关系。

(4) 营销渠道:营销渠道就是将产品或服务推向客户的一整套相互依存的组织。在知识经济时代,产品和服务的知识含量不断提高,营销渠道和传统的渠道有很大的不同,其中包含很多知识传播和交流成分。对于非营利性机构来说,这类资本乃是与其他方面的关系,因此称为关系资本会更恰当一些。

4. 知识产权资本

所谓产权,就是社会制度所授予特定个人的某种特定的权威。这种权威规定特定个人或组织从财产(产品)的多种使用方式中有权选择一种方式加以使用,以获得相应的效用。

完整的产权包括一组权利,即使用权、处置权、收益权和转让权。

(1) 使用权。在法律许可范围内,以自己喜欢并且是自然规律许可的方式使用该财

产（产品）。

（2）处置权。也就是对财产的处理权利，包括对财产在物质形态上进行任何改变，直到抛弃重置。

（3）收益权。这是通过产品获得收益的权利，包括直接通过使用财产本身创造收益，或由团体协约关系从他人那里获取收益。

（4）转让权。将财产通过契约出售或出租，获取财产权利让渡收益，或租金收益。

产权结构是由多个层面组成的，要想产权不受侵犯，必须从每个层面加以保护。国家内部建立了法律制度，包括了产权制度。社会建立了产权保护法律，强制推行，使其成为社会准则。

现在人们已经认识到，知识在一定条件下，具有产权价值。知识产权乃是对人类智慧结晶这一无形财产所享有的权利。1967 年，《建立世界知识产权组织公约》把知识产权定义为：在工业、科学、文学或艺术领域里的智力活动产生的所有权利。

3.3 知识管理的任务与原则

这一节将介绍知识管理的任务和知识管理的原则。

3.3.1 知识管理中的主体与客体

对知识管理的主体和客体的识别有助于人们对知识管理的任务和目标的确定。知识管理的主体是人和由人组成的群体；客体则是人们所认识或处理的对象，包括自然界、社会、个人或组织的行为和思想。

知识管理的主体可以分为个人、组织（包括企业和机构等）、国家乃至全人类社会这样几个层次。

对个人来说，要想适应当前的社会对自己的要求，就需要不断地充实自己的知识。学校教育虽然提供了比较系统的知识，但是由于知识的更新和发展速度越来越快，人们不仅需要通过工作中的边干边学来得到经验和知识，还需要不断有计划地补充自己的知识。不但个人进行自学、总结经验和进行思考，还需要与别人交流。因此需要对自己所掌握和所需要学习知识有一个筹划，要不断进行梳理，也就是说要对个人的知识进行管理。

在组织层面，对一个企业来说，为了获取经济和社会效益和尽到自己的社会责任，需要不断地将知识运用到经营管理、生产和服务的各环节，并且不断地进行产品、生产过程和经营模式的创新。创造新的知识，这是当前知识管理的关注重点。而一些非营利性的组织，如图书馆、文化组织等，需要不断地向人民提供各类知识。至于学校和研究

院所，更是需要不断地获取知识、传播知识和创新知识。

在国家层面，在不断变幻的国际形势下怎样制定和执行政策，才能使国家可持续发展，也需要制定发展战略和创新政策的知识，需要实施战略和执行政策的知识，需要应对突发事件的知识，因而需要进行宏观的知识管理。

当今世界各国在面对威胁整个人类的一些因素和事件时，也在联合起来共同寻找解决问题的知识。

知识管理的客体所包含的内容十分广泛，随着人类活动空间的日益扩大，人类活动的方式日益增多，人们关注的知识内容的增长极为迅速。这从当前知识门类的迅速增加和每门学科内容的迅速膨胀就可以理解到这一点。

如果想把人类关注的知识领域完备地列举出来是极其困难的。但是综合起来看，知识管理的客体可以按照知识内容和知识过程两个方面来划分，这是与 1.5 节所说的知识的对象与过程两重属性相对应的。

以知识内容为知识管理的客体，是从静态的角度来看待知识的。需要掌握知识本身的基本性质（如知识的可表示性、可传递性、可处理性、可共享性、可复用性等）以及知识在何种条件下成为公共物品，在何种条件下成为商品，以及知识的价值在哪里等问题。

以知识过程为知识管理的客体，是从动态的角度来看待知识的，需要掌握知识的运动过程，包括知识的获取、知识的存储、知识的处理以及新知识的生成过程。从技术侧面探讨的是知识怎样通过技术网络或社会网络进行传递、转换的过程，而从经济侧面探讨的是知识怎样通过价值链进行增值。

知识管理的对象是知识内容与知识过程的辩证统一，而整个知识管理则是知识管理的主体与知识管理的客体的辩证统一。

我们之所以强调认识和区分知识管理的主体与客体，就是要强调知识管理是以人为本，是人在掌握和控制支配知识，而不是被知识所控制和支配。这样才能避免陷入技术至上和对知识盲目崇拜等误区。

3.3.2 知识管理的任务

知识管理的任务包括：

（1）构建组织的知识战略以及知识经营管理的策略，并组织实施。

对一个组织来说，要想充分发挥知识的作用，需要从宏观上制定组织的知识战略，而知识战略又是要符合组织总的发展战略的。对国家来说，知识战略是国家可持续发展的总战略的有机组成部分。对一个企业来说，知识战略需要保证企业总战略的实施，特别是企业创新战略的实现。

(2) 掌握本组织为了生存和发展对知识的需求,动员内部的各种力量去获取各种知识资源,加以组织和处理,使其有效地运用到工作实践中去。

当前任何企业或组织,要想在激烈的竞争中立于不败之地,需要不断进行变革以适应环境的变化,这就需要不断获取新的知识,既可从外面获取,也可在内部挖掘。也可以从合作伙伴甚至于竞争对手那里获得有用的知识。同时还要在现有知识不足的情况下进行知识创新。

(3) 鼓励和组织员工学习和积累知识,激励员工参加知识创新活动,营造学习与创新的环境和氛围。

由于社会经济环境的不断变化,只有经常学习才能适应竞争的需要。在组织内部,要使学习蔚然成风,使自己的单位成为学习型组织。还要通过一切措施鼓励员工投入创新活动之中,发现和挖掘员工的创新潜力。

(4) 建立知识流通和知识共享的机制和条件。

知识只有流通才能不断产生价值。应该建立有利于知识流通的体制,创造知识流通的技术条件和制度条件。

(5) 建立对知识的存量和流量进行检测和评估的制度和方法。

只有对知识的积累和流通,以及对知识创新活动与成果不断地进行检测和评估,才能对本组织的知识现状和成果做到心中有数,才能发现自己在知识工作方面的不足和明确今后的方向。

(6) 保存、保护和经营好知识资产与知识资本。

对现有的知识资产,无论是有形的还是无形的,都要加以保存,建立知识库。要经营知识资本,使其增值,同时还要注意对知识产权的保护。

所有上面这些任务和活动都是与组织的业务活动紧密联系在一起的,也就是说,知识管理和组织的经营管理紧密相连。

3.3.3 知识管理的原则

要想有效地实施知识管理,应该遵循一些从实践经验中总结出来的原则,这些原则是符合前面提到过的知识特点的。下面列举其中的一些主要原则供参考。

(1) 知识管理必须以人为本。因为知识是人创造的,又是被人所运用的,所以对知识的管理必须以人为本。

(2) 知识管理必须兼顾显性知识和隐性知识。因为知识在发挥作用时,总是两类知识同时在运用。由于长期以来人们对于隐性知识的作用缺乏自觉的认识,因此在知识管理过程中需要特别对隐性知识加以关注。

(3) 按照知识的生命周期进行管理。知识也是有生命周期的,应该按照知识的生成、

发展、成熟、衰退和消亡过程，遵循知识的演化规律来处理知识。

(4) 知识应该在传播和交流中创造价值：只有在交流过程中做到知识共享，才能使知识发挥更大的作用。

(5) 知识的应用和创新是全体员工的事：只有全体员工都关心知识应用和创新，才能使知识发挥更大的作用。

(6) 应该有专门的组织和人员从事知识管理工作：他们的职责是组织和管理知识的应用和创新工作。高层管理人员也需要有专人来加以领导。

(7) 应该营造有利于知识管理的企业文化：知识的应用与创新不仅是单纯的技术问题，而且是与组织的制度建设、尤其是企业文化息息相关的问题，因此需要培育有创新精神的企业文化。

3.4 知识系统与知识系统工程

这一节先介绍知识系统的概念，然后讲述如何利用系统工程思想与方法来研究和实施知识管理。

3.4.1 知识系统

人类的各种知识是相互关联的，形成了不断发展演化的知识系统。知识的传播、运用、创新及其管理具有一系列的系统特征。

我们不妨从系统的角度来研究知识。从系统学科的基本观点看来，系统是由许多要素组成的整体；要素之间，要素与整体之间，整体与环境之间存在着有机联系；由于这类联系，使得系统涌现出整体功能，即系统功能；系统存在于一定的环境之中；系统是有层次的；系统的存在与发展必须适应环境。下面就从系统的这几个基本属性来审视知识系统。

对一个组织现有的知识系统来说，系统中的要素如下：

- (1) 存在于文件、手册和图纸中的知识；
- (2) 存在于人的头脑中的知识；
- (3) 已经凝聚在工作过程、经营管理制度和方法等之中的知识；
- (4) 嵌入（外化）在产品或者服务中的知识。

由于知识的高度抽象性，人们更容易从知识的拥有者或者知识的载体这样一些看得见的实物、人员或者组织机构来辨识知识的存在。所以人们看到的乃是：保存知识的书刊、文件、计算机的数据库、光盘、磁盘等载体以及有知识的人员，还有与知识密切相关的组织。但是他们所保有或承载的知识本身、特别是人所掌握的知识，才是构成知识

系统的基本要素。

系统是存在于一定环境之中的。所谓系统的环境，是指与系统发生相互影响、相互作用，但又不属于这个系统的所有事物的总和。一个企业的知识系统的环境，就是这个企业所在的地理环境、外界的社会经济和人员等方面各种因素的总和。

按照系统思维方式，可以从系统的结构和功能两方面来对知识系统加以考察。

系统的结构就是系统中的要素之间相互关系的总和，它形成了系统内部相对稳定的组织形式和结合形式，形成系统的整体。要素的相互作用对要素起到了联系、限制和约束作用，从而形成了系统某种稳定的结构模式。例如，物理学的知识中，不但力学、热学、电磁学等各自领域内的现象和规律之间存在着一定的联系，而且力学、热学、电磁学等领域之间也存在一定的联系，这样才能构成该学科整体的知识。

知识系统是以网络结构形式存在和运行的。网络的一部分是有形的，如信息网；也有一部分是无形的，如人与人之间的关系网，合作默契关系。

系统的功能乃是系统在与环境相互作用时的总体行为、特性和作用的总称。人造的系统是有目的的，系统功能就是为达到目的的行为和能力。

知识系统整体功能应支持知识主体的存在和发展（对国家来说是提高创新能力，增强综合国力，对企业来说是提高企业的竞争力和活力，对个人则是提高他的专业水平和素质）。

知识系统的功能应该包括：

- (1) 高效率高效能地获取和组织知识。
- (2) 能够有效地保存和保护知识。
- (3) 能适时将知识传播到适当的地方给适当的人。
- (4) 能高效率和高效能地开发新产品或新的服务。
- (5) 能按市场规律经营和管理知识资产。
- (6) 能营造和加强有利于知识生成、转移、使用的组织文化。

为了实现上述功能，上面提到的各类知识又是与组织的业务工作过程或者人员紧密联系着的，起到指引、支持、改善业务活动的作用。

知识管理就是对知识系统的管理，在 3.3 节谈到的知识管理的任务，就是围绕着知识系统的功能提出的。我们在本章 3.1 节曾经说过，知识处理和知识管理是相互联系、紧密结合的两类活动，因此知识系统中包含两类子系统：一类是知识处理系统，如企业中的研究开发和设计系统，另一类则是对知识处理进行管理的知识管理系统。这两类系统形成了一个两层结构。

由于组织中的工作过程或人员也是以系统形式存在的，因此知识系统与业务工作过程系统或人员系统集将成为一个更大的系统。

知识系统不能仅仅理解为以技术网络为主体构成的系统，而是以人所构成的社会网络和技术网络、知识内容网络所集成的复杂系统。它不像一般的实物系统或者传统的信息系统，一旦建成，就能够按照既定的目标和要求工作。由于知识的生成是人的复杂脑力劳动以及人与人之间互动的产物，而每个人以及每个团队、组织都有一定的自主性，所处的环境又有一定的不确定性，所以说知识系统是一类复杂的自组织系统，新知识的产生乃是系统的“涌现”产物。尽管具体的以计算机和网络为核心的支持知识管理的技术系统的构建与运行可以使用信息技术来完成，但是对于广义理解的以人为本的知识系统来说，知识系统是一类人-机结合而以人为主的系统，它的构建与运行是一种社会活动。

3.4.2 知识系统工程

我们希望用系统工程的思想和方法来建立实际应用的知识系统。这首先要按照知识系统的特点，建立一套针对知识系统的系统工程原则和方法。

系统工程这个名词，可以有两种解释：一种是指一项影响面广、涉及多种因素而综合性较强的工程项目或者建设事业。这就像人们常常说的：企业信息化是一项系统工程，预防流行性感冒是一项系统工程，等等。另一种解释则是指一个学科，也就是研究怎样实施系统工程项目的技术学科。本书所提到的系统工程，则是指后一种解释。

按照中国系统工程学科开创者钱学森和许国志两位先驱者的说法，系统工程是关于组织、管理的技术。它是涉及工程实践的学科。现在人们对于工程这个名词的理解，已经不限于技术领域中的那些建设项目了，一些社会经济乃至科技文化建设都可以称为一项工程。

系统工程学科的核心内容是系统工程方法论，它所研究的是在工程实践中解决系统问题所用的思考问题的独特方法，以及分析工具和工程实践步骤。它研究的是一般性的指导原则，可以适用于各个领域，而应用到每一个具体领域（这个领域不一定仅限于一个专业，而是可以包含许多同类专业）时，还要结合该领域的特点与特殊要求，因而形成了各领域的系统工程专业学科。

知识需要管理，而知识的生产、传播与利用又形成了一个系统，那么是否可以考虑把知识系统的研究提到系统工程的高度，从系统工程的一个新的学科分支：知识系统工程的角度来进行研究呢？

我们曾经建议创建一门新的学科：知识系统工程，作为用系统工程语言和思路来研究知识管理的学科，并把它定义为：

“知识系统工程是对知识进行组织管理的技术。”

系统工程作为一门学科，产生于 20 世纪 40 年代，开始处理的对象主要是技术设备和系统，后来发展到应用于社会经济、生态等更广泛的领域之中。原来研究物流、能流

及信息流,现在则还要研究知识流。管理知识的系统和管理物流、资金流的系统有很大的差别,应该根据各自的特点,建立自己的思路和方法。

由于知识系统的高度复杂性与抽象性,对于系统的研究常有无从下手的困难。在实物生产的筹划和实现过程中,只要人们做出符合客观实际的计划和设计,并有充分的资源保证,大多数都能达到预期。但是知识生产却不然,人们无法像实物生产那样规划和预期,有时候你想得到的不一定得到,而在意外的情形下却会得到意想不到的结果。正因为知识的上述特性,使得人们在研究知识系统时,无法照搬实物系统的研究方法,需要探寻新的途径。

这里要特别说明的是,知识系统工程中所研究讨论的知识系统,其含义与内容要比人工智能中知识工程的研究对象——基于知识的系统(Knowledge-based System)广泛得多,因为知识系统工程研究的对象不仅包括信息技术工具(硬件和软件),而且还包括知识系统的主体——知识工作者在内,不但涉及技术问题而且涉及经济、法律和文化等问题。基于知识的系统可以被看作是知识系统中的一种工具,一种子系统。关于这一点,我们在后面还要提到。

由于关于知识系统的研究还处在发展的初期,要想建立一个基于知识系统工程学科的理论框架还需要做许多工作,这里试图从下面三方面来构筑一些简单的架构:

- (1) 知识系统的体系结构;
- (2) 知识系统的运行过程分析;
- (3) 知识系统工程项目的开发。

其中,知识系统的体系结构从静态角度描述了系统的组成结构,知识系统的运用过程从动态角度分析了系统的工作过程。

现在先研究知识系统的各种体系结构。

在本书第2章曾经说过,体系的体系结构用来表述系统某一方面特点的结构方式和模块组成,它侧重于原则、侧重方法而不具体规定技术和业务细节。它着眼于为实现系统功能而做的原则性的安排。这种安排着眼于各部分的关系应该怎样处理,使得系统在整体上结构合理,有关部分各得其所。

由于系统工程所研究的对象都是人造的系统,在建立这样的系统时,先得有一个从全局的高度来考虑系统究竟应该如何构成,以满足要求。也就是说,先要搭建系统的架构。体系的体系结构正是从不同的视角来描述系统架构的。

建立系统体系结构的工作叫做“建构(Architecting)”,这是系统工程的基本任务。建构应该有科学基础,但由于每一个具体的系统都带有自己的独特性质,不能千篇一律地用同一种方法,所以说建构工作既是科学,又是艺术。正如中国书法讲究“间架结构”,每个汉字同样是那些笔画,但是布局不同,写出来的神韵就不同。

这里尝试着提出研究企业知识系统的一种体系结构的设想,这个体系可以用图 3.1 所示的轮形结构来表示。

这个总的体系结构包括人员体系结构、组织体系结构、技术体系结构、经营体系结构和文化体系结构五个方面。

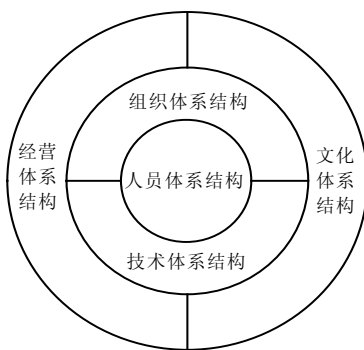


图 3.1 企业知识系统的体系结构

下面对各类体系结构略作解释:

(1) 组织体系结构,它说明知识系统中的有关组织机构及其相互之间的关系,它与人员体系结构关系也比较紧密。

(2) 人员体系结构,它说明在知识系统中有关人员的组成以及相互之间的关系。由于我们认为知识管理是以人为本的,所以在图 3.1 中,人员居于轮式结构的中心。

(3) 技术体系结构,它说明知识系统中的技术工具和技术方法的组成及相互间的关系。

(4) 经营体系结构,指知识系统在实际应用时所涉及的经济、管理等方面的运作规则和度量评测方法等。

(5) 文化体系结构,这涉及与知识系统有关的组织价值观、愿景、目标、信念,以及人际关系等文化环境因素和精神因素。

这些体系结构从不同的侧面描述了知识系统组成要素之间的关系,以及它们的作用。本章下面几节的内容将对每一种体系结构中的主体及其任务做一些介绍。

我们再来研究知识系统中的过程。人们在工作过程中,要通过获取、处理、传播和应用知识的一系列过程,有时候还会生成新知识的过程,才能有效地使知识发挥作用。这些过程的总体形成了知识过程,或者说知识流程。在过程中知识从一个环节流到另一个环节,有的知识则外化到产品、服务以及组织形式、规章制度乃至企业文化中去了。

知识过程应该再划分为知识运作过程和知识管理过程两大类。就像在实物生产过程中,有生产加工过程(如机械加工过程和化工技术过程等),也有生产管理过程一样。知识运作过程是一线工作人员在进行业务工作时获取、处理与应用知识的过程,它与业务

过程紧密相连。知识运作过程与业务过程有一致的地方，特别是在研究开发等创新过程中；但也有差异的地方。例如，产品设计完了之后，交付设计是业务过程，而写总结报告、积累经验却是知识过程。至于知识管理过程，其含义与内容要广泛一些，从结合企业战略目标确定知识战略，到提供知识运作过程的物质条件（信息工具等）与组织和文化条件，都应该是知识管理过程的内容。

这样一来，两种过程就形成了一个两层结构，底层是与业务过程密切结合的知识运作过程，上层是知识管理过程，它对知识运作过程进行组织和管理。

在当前知识经济发展的萌芽期，知识运作过程和知识管理过程有相当一部分是混在一起的，但从功能上还是应该把它们区分开，这样容易界定知识管理的任务和内容，以及人员与学科的分工（因为知识运作过程的相当一部分是人工智能中的知识工程的研究对象）。下面将利用系统分析思想来把它们综合集成起来加以研究。

在知识运作及其管理的过程中包含：

- (1) 知识需求的认定。根据业务内容与特点，以及组织的知识战略，确定所需的知识。
- (2) 知识的获取。先要检查本组织已经具备的知识中，是否有所需的一部分知识，再决定还要哪些新的知识。新知识的获得有两种可能，一种是通过在组织内部寻找，一种是从外面得到所需的新的信息与知识。
- (3) 知识的组织。对知识加以解释、分析、编辑、索引、过滤、综合、链接和存档，并准备随时检出应用。
- (4) 知识的转化和创新。如果缺少的知识无法通过上述步骤获得，则需要通过在显性知识与隐性知识的转化及个人知识与组织知识的转化过程中进行知识创新，生成新的知识。但生成的知识仍旧是未经检验的，也需要验证。
- (5) 知识的传递。这个传递指传递给有关的使用者，在传递过程中，接受者要加以检验和吸收，然后在实际中加以应用。
- (6) 知识的应用和创造价值。这里所谓创造价值，除了知识本身是用来增值外，还可以转让。

关于知识系统的开发，将在本书第 12 章讨论。

3.5 知识组织和知识工作者

这一节首先探讨知识管理中的组织和人员问题。

3.5.1 知识组织

知识组织涉及知识的产生、传播、有效利用的组织原则与形式。这里涉及的组织问

题有两个：一个是专门进行或管理知识工作的组织体系（可称为职能性知识组织）应该如何建立的问题；另一个是整个企业、院所的组织体系由于知识经济的发展，为使知识能够更好地发挥作用而需要重构的问题。下面依次加以讨论。

职能性知识管理沿袭着原来的组织结构，目前具备两种形式：一种是经常性管理的组织，一种是按项目管理的组织。

经常性管理是把知识工作的管理纳入经常工作。这些知识工作是企业在日常运行时有关职能部门的本职工作。知识管理的组织已经纳入现在已有的组织形式中，如当前大多数组织还是科层式的，在相应的层次上应有专门的机构，如技术部、设计处和技术监督部等。随着企业的扁平化与网络式组织的出现，专门机构的位置也会相应地改变。

按项目管理是为了产品开发、技术更新改造等创新需要而组织专门的课题组、攻关组，项目结束时组织也就自然不存在了。

两种组织形式各有利弊，在经常性管理的现有组织体制中，由于日常工作接触，易于了解生产或业务实际，易于与现场人员合作沟通，但常常因为阶层制的官僚式体制阻碍或延缓了知识的流转。按项目管理的课题组任务明确，精力集中，知识易于沟通。但有时无利于现场了解和推广成果。

由于知识管理的需求日益增加，原来的职能性知识组织已不能满足要求，目前国外也在探索新的组织形式。

各类知识组织必须能实现知识的共享；使企业内外的知识（包括显性和隐性知识）能够被识别和获取；使组织的新知识得以生成，知识创新得以实现；各种形式的知识资产能够加以控制并随时提供应用；还要能为知识工作者的协作创造条件。

对一个企业来说，知识组织可以分高层、中层和基层三个层次。

（1）知识管理的高层组织是企业最高领导层面上全面负责知识管理的机构，如专门的委员会和领导小组等。现在一些研究院所、高等院校中设立的学术委员会，就是这类组织。总负责人一般就是组织的最高领导，而后面将要提到的知识总管则是这类机构的执行负责人。这类组织的任务是制定企业在知识方面的战略，决定相应的组织、人员和投资等方面的安排。

（2）具体执行知识管理任务的是一些中层的管理组织（知识组织的中层），如前面说到的技术部、设计科、研究开发中心，等等。今后还会按需要成立知识中心、知识网络、转化中心和孵化中心等。

知识中心实际上就是基本的知识资源库，它统一获取、储存、分发各类显性知识（包括书刊、文档和网上知识资源）。

（3）基层的知识组织就是从事具体知识工作的团队。现代化的企业基层组织，有向团队化发展的趋向。这在知识型工作或知识型组织中更能显出它的优越性。试以产品开

发设计工作为例来看它的特点。

传统的产品开发工作是以一种串行的方式进行的。首先是设计部门技术人员进行产品内外设计；然后将图纸交给工艺部门进行产品制造工艺设计；再交给生产技术准备部门进行工艺装备、测试仪器等生产技术准备；采购部门进行采购；一切都准备完了以后生产部门开始试生产。每一个部门在工作中很少与其他部门进行沟通，往往只是站在本部门本专业立场上考虑问题，这样常常造成设计修改大循环，影响产品上市时间，不利于竞争。如果采用跨部门团队形式进行产品开发就没有这些缺点，可以大大提高开发的成效。产品开发团队由来自工程设计、工艺设计、制造、后勤、材料和质量保证等各部门的代表组成，根据需要可聘请其他专家，还可公开邀请用户、供应商参加。这样就会使设计工作更有整体性、合理性和可行性。团队工作的实质是从原来面向功能的工作模式，转向面向过程、面向产品、面向结果以及面向顾客的工作模式。

团队工作方式的技术支持是计算机信息技术。由于信息技术的发展，中层管理者作为信息传递以及监督的作用开始减弱。将基层员工组成各种面向产品或过程的团队，并适当的授权，就能使组织扁平化。

除了正式组织之外，一些非正式组织在知识的管理中也会发挥不可忽视的作用。这类组织是由一些知识工作者按照共同的兴趣或者共同的观点和见解而形成的。其中有一些组织还是比较正式的，如专业学会、研究会和讨论班等；还有一些组织则是没有固定形式的，知识的作者会经常聚集在一起进行知识交流。

3.5.2 知识工作和知识工作者

在知识系统中，知识的创造和应用都是由人来进行的，人是知识系统中的重要主体。他们在系统中究竟各自从事什么工作呢？他们的工作特点又是什么呢？这就是下面所要探讨的问题。

知识工作是经常生成新知识的工作，它的核心任务是思考，它需要脑力进行工作，成果是信息与知识，工作常常是无法按直线顺序完成的。

有人认为知识工作是一类职业，有人认为这是一类工作的性质，也有人认为是一类工作的特点，本书取第一种理解，把它看作是一类带有特征的职业。

知识工作具有下面四个特征（按美国劳动统计部门的规定）：

- （1）知识工作是由众所公认的可言传的知识支持的；
- （2）这些知识可在书本中找到；
- （3）掌握知识的人必须经过官方或学院认定；
- （4）职业必须有专业的行为规范。

所谓“知识工作者”，就是运用知识进行工作的人，是用脑力劳动的专业人士，他们

所从事的就是“知识工作”。知识工作者主要是用智力、用头脑来工作，而不是用体力和手来工作的。知识工作者自己不能单独生产任何有形产品，他们提供的只是信息、创意和知识。

当前，任何组织中的高层、中层和基层技术、管理人员都与知识工作有关。高层人员一般是起战略决策、引导方向和进行评价的作用。基层知识工作人员亲身接触实际，起到获取、传播、积累和生成知识的作用。中层人员则是承上启下，在领导的设想、意图与实际工作之间起桥梁作用。

其中，基层人员包括规划人员、设计和制造工程师、市场研究人员、科技信息工作者。随着知识经济的发展，这一部分人的数量会日益增多。

由于信息是知识的载体，信息工作者所获取、处理与传递的信息中一部分是包含知识的。但是信息工作者的任务不涉及知识的直接应用和新知识的生成，知识应用和生成的工作是由知识工作者完成的。体力劳动者也不是一点也不需要知识，无论是开动机床还是驾驶载重车也都需要一定的专业知识，但他们不像知识工作者那样频繁地与知识打交道。随着自动化与数字化技术的发展，操作数控机床需要的知识量也会增加，操作者的身份也会逐渐发生变化。

如果再把知识工作者细分一下，可以把它分成知识运作工作（或称知识处理工作）者和知识管理工作者两类。

像进行新产品研制、产品或者工程设计、生产计划与调度、故障诊断这一类的工作都属于知识运作（知识处理）工作，因为在工作中要获取、加工和应用知识，有时还要生成新知识。从事这类知识工作的人，可以称为知识运作（知识处理）工作者，通常所谓的专家或专业人员（包括高级技工）就属于这一类。这类人可以统称为知识员工。

另一类知识工作是专门从事知识的组织与管理，即知识管理工作，如对一个组织的知识进行规划、筛选、保存和传播，以及对知识工作者进行组织和管理（因为对知识工作者的管理就是对他们所具备的知识、特别是隐性知识的管理）。从事这类知识工作的人，可以称为知识管理工作者，3.5.3节将介绍这一类人员。其实在知识密集型企业中的人事部门中的管理人员，就承担了一部分这类工作，因为他们要管理知识工作者。

目前，我们还无法非常明确地把这两类知识工作划分开，因为知识运作工作中也还连带着一些知识的组织管理，如知识的获取、筛选、存储和传递。这就有点像在工业经济的初期，在小的工厂、作坊中也很难划分生产加工和简单的生产组织管理一样。

知识管理工作有的是从信息管理工作转化而来的，其中主要的是可编码知识资源的管理，但也不乏知识的挖掘与新知识的生成；有的是对知识工作者的人事管理，而更重要的却是如何规划、设计和组织知识的获取、存储、传播和有效地利用，以及新知识的产生。

在知识管理人员当中，最重要的是知识总管（CKO）。知识总管是一个企业内部专

门负责知识管理的高层管理人员，是随着知识管理日趋重要而在企业中设立的像技术总管、财务总管、信息总管一样的高级职位。他的工作内容大致有下列五个方面：

(1) 营造一个有利于知识应用和发展的环境，包括构筑知识获取、共享和应用的基础设施，形成有利于知识发展的文化氛围。

(2) 按照企业的当前与今后的知识需求，组织引进各类知识，进行知识的对外交流。

(3) 促进组织内部的知识交流与共享。

(4) 更重要的是指引知识创新的方向，集成各类知识，推动创新，包括技术创新，管理创新和市场创新等。

(5) 规划与组织企业内的学习和培训等掌握新知识的活动。

知识总管不仅是一位技术专家，而且应该是管理专家，特别是战略专家。因为他不但要能领导知识工作中的技术工作，如信息采集和管理，而且必须有战略眼光，能够洞察未来的知识发展和知识需求。

知识总管需要具备多方面的知识和能力。从宏观视角来看，应该具有经营知识、政治和社会知识、风险分析知识和能力、影响力、领导能力、创造力和表达能力的知识和能力。

在规划和实施方面，知识总管应该具有组织开发、信息战略、财务计划、沟通能力、创新工作、风险管理、灵活性与开放性、跨领域管理、帮助个人进行自我管理、释放他人全部潜能的知识和能力。

除了知识总管外，现在各国有关知识运作和知识管理人员的职业称呼可以说是五花八门。例如，知识分析师、知识工程师、知识编辑、知识导航员、知识架构师、知识经纪人、知识资产管理、创新经理/官员、知识集成师、知识领导人和内部沟通经理，等等。

最近还出现了像知识项目经理、知识部门经理等知识管理人员。知识项目经理的职责是企业内部的知识项目进行筹划，作可行性研究，组织队伍，安排计划，调度资源，利用信息技术环境营造知识系统架构，组织设计和实施，保证经费及时到位，监督工作的进行，解决一些临时发生的问题，进行验收总结，等等。由于知识项目不同于一般的工程项目，涉及物的因素少而涉及人的因素多，因此对知识项目经理的要求需要考虑其处理人事工作的能力。有时需要设置知识部门经理，其实现在许多企业中的研发部门、设计部门等的领导，实际上已经承担了知识部门经理的工作。但是为了突出知识的独特作用，设置这样一个岗位也还是必要的。

还有一种专门岗位，就是知识分析师（知识分析员）。他很像信息系统开发工作中的系统分析师（系统分析员），但是重点在知识系统。对于一个有关知识运用和创新的开发项目，知识分析师的作用是很重要的，他起着知识系统的用户和系统开发人员之间的桥梁作用，是系统开发过程中的关键人物。他要负责了解用户的想法和需求，运用自己对

知识内容和处理知识工具的专业理论和技术,以及系统的开发方法和步骤,形成明确的知识系统的概念。知识分析师应该具备两个方面的知识和能力,一方面是技术方面的,包括有关专业的业务和知识管理技术以及系统开发技术方面的知识和能力,另一方面是处理人际关系方面的知识和能力。

此外还有管理知识资源、管理知识工具的技术人员。

3.5.3 知识工作者的特点和对知识工作者的管理

知识工作者(或称知识型员工)由于其从事的工作之特殊性,一般具有下列特点。

(1) 知识工作者一般都具有较高的学历和专门的科学文化教育背景,像一些高新技术产业以及技术性较强的部门,从业人员中高学历的比重都是很大的,更不用说像高等院校、科研院所那些知识密集的单位了。知识工作者所具备的某一方面的专业知识,常常要比他的顶头上司还宽广和深厚。

(2) 知识工作者的日常工作是获取知识、加工处理知识和应用知识,有的还要创造新的知识,而世界上的知识总是在不断地更新,所以尽管他们已经具备相当水平的知识,但仍旧需要不断地学习,获取新的信息和知识,跟上时代的发展。

(3) 知识工作具有一定的自主性或者说自治性,如像一些研究、规划、设计等创新性的工作,对个人独立思考的需求要比集体研究讨论更多,因此知识工作者常常希望能有一定的授权和独立工作条件。

(4) 知识工作由于有创新的要求,因此知识工作者常常希望减少一些限制和约束。为了适应环境变化,无论组织方式还是运行机制,希望有一定的灵活性。知识工作者对于一些官僚式、命令式的领导方式是难以适应的。

(5) 知识创新工作有一定风险,所以知识工作者希望能有一个宽容的环境,允许失败。

(6) 由于知识工作者的知识主要保存在他的头脑之中,而知识工作者之间的协作以及组织知识的形成都需要进行知识的交流与共享。知识工作者在贡献自己的知识时,一方面会顾虑到一旦知识共享之后,会丧失他本人的优势甚至丢掉工作,因而常有保守知识的倾向;另一方面他有一种社会责任感,希望对社会有所贡献,同时会因在知识方面有所创新并得到别人的认可而有一种自我实现的成就感,究竟哪种因素占上风,就要看组织对知识创新的重视程度、激励方式以及组织的文化氛围了。

(7) 知识工作者除了重视成果的物质报酬外,还十分重视精神报酬。

(8) 知识工作者,特别是专业性很强的知识工作者,由于经常单独工作,使得其中的一些人不善于与别人合作,还会有狭隘的专业偏见。也就是说,知识管理领域的专业人员和以信息管理为手段进行知识管理的人,容易忽视人和组织因素;从事人力资源管理的人,容易忽视技术因素。

对知识工作者的管理必须以人为本。所谓以人为本,就是尊重人的不同层次的需要,把人看作是追求自我实现和能够自我管理的“社会人”,而不是仅仅追逐物质利益的“经济人”。因此要以开发知识工作者的潜能为主要目标。

以人为本的管理,首先要确立知识工作者在管理过程中的主导地位,不但要激发和调动他们在技术知识工作方面的主动性和创造性,还要激发和调动他们在管理过程中的自觉性、主动性和创造性。使他们觉得自己也在参加管理而不是仅仅被管理。

以人为本的管理,还必须注意实施以自我管理为主的方式,外界主要是诱导和启发,减少单纯命令式的管理。使得知识工作者心情舒畅,得到全面发展。

在知识工作者的管理上,必须注意:

(1) 由于知识工作的自治性或者说自主性较强,所以在组织形式、工作安排上应该充分授予自主权,在领导方式上要充分尊重知识工作的独立思考和自由发表意见。

(2) 在知识工作者的集体中,领导与被领导的界限变得模糊起来,因为从知识能力来看,下属可能比上级具备更多、更深入的知识。所以知识性组织的领导就应该像一个乐队的指挥一样,在某一种乐器的演奏和发挥上不可能比专业演奏者更强,而指挥本人的任务是协调整个乐队的演奏,使每种乐器的演奏都能在服从整个乐曲要求的前提下,由演奏者来发挥。

(3) 为了实现知识交流与共享,以促进知识工作者之间的协作和形成组织知识,必须鼓励知识工作者积极贡献自己的知识。另一方面,知识工作者具有较高的流动意愿,希望能够在更能发挥自己才能的环境中工作,怎样吸引人才和留住人才,就成为知识管理中的重要任务。

(4) 对知识工作者个人来说,也要进行个人的知识管理,也就是对个人的知识进行获取、组织、存储、利用和创新的管理,管理过程是在不断明确自己的知识需求的基础上,有效地识别和获取、整理和存储、集成和开发自己的知识的过程。一些知识工作者(主要是知识处理工作者,如研究开发和设计人员)一方面要管理自己的知识,另一方面又要作为被管理的对象,接受知识管理者(如他们的上级和知识总管)的管理。

有的学者认为,知识管理是从个人开始的。在这个基础上建立团队和组织的知识管理,然后再扩展为组织之间或更大范围的知识管理。事实确实如此,无论是个体知识工作者(如理论研究人员、律师和作家等),还是团队成员,都必须从明确自己的知识需求开始,然后收集和整理,由自己或者与同事共同运用这些知识,也可能在此基础上创造新的知识。可以说个人的知识管理乃是组织知识管理的基础。

个人知识管理不仅是就业或创业的人才需要,在校的学生其实也需要个人知识管理,只是与在职人员的目标和要求不同而已。

3.6 知识运作与管理的技术设施

本书前面已经多次说到,知识管理是以人为本的,人和组织是知识系统中的活动主体,但是一些信息工具还是起了辅助与支撑的作用。知识存在于两种载体之中:一种是生命载体,也就是掌握知识的人;一种是物质载体,如书籍、期刊、磁性和光学媒体(磁盘、光盘)等。在知识获取、传播、使用和产生过程中,人与人之间、人与物质载体之间,以及物质载体彼此之间要进行知识交换,知识也还要存储,有时候还需要将信息转化为知识。知识的运作与处理需要技术设施和技术方法。

3.6.1 知识运作与管理的技术工具

信息化工具与信息系统在知识运作和管理中有着极其重要的作用。计算机的高速处理能力加快了人们获取和处理知识的速度和深度;大容量的存储和方便的检索使得人们很容易积累数量巨大、门类繁多和形式多样的知识;网络化消除了空间限制,全球化的网络(目前的互联网和今后的第二代互联网)成为一个庞大的知识来源,虚拟现实技术使人们能像身临其境一样模仿一些难以具体实现的场景,提高了学习和实验能力,凡此种种,都对人类的知识与智能的发展起到了促进作用。因此需要研究信息技术如何在知识运作和管理中发挥更大作用。

在知识获取过程中,人们既要从组织内部、也要从组织外部获取有用的信息和知识。各种图书、期刊和公报等都是信息与知识的来源。现在的互联网是一个特大的知识库,可以从网上获得国内外有关的信息和知识。

对于显性知识,可以通过网络,使用搜索引擎和浏览器去查找有关的知识,也可以先获取有关数据(这可以通过组织中已有的信息系统中的数据库或数据仓库得到),再通过数据挖掘、知识发现等技术获得相应的知识。

对于隐性知识,由于它难以用语言文字表述,很难通过搜索,找到它在什么地方。因此,有效的办法是能够找到谁具有所需要的知识。这样一来,对知识的搜索变成了对掌握知识的人的查询。现在有一些编制专家网络图的方法,用来帮助人们找到掌握相应隐性知识的人。

获取的知识要加以过滤,以保证其正确性与一致性,要加以分类以利于以后查找。

对获得的知识需要用适当的方式加以表达,还要加以编纂,以利于存储、共享和处理,并便于传播。编纂的原则是明确它的服务目的和用途以及适用性。对显性知识,可以编码输入数据库;其中对文本形式的知识还有如何进行编写“内容摘要”的问题。对隐性知识,由于难以表述,可以考虑用含有隐性知识内容的案例、图形和视像等编辑到存储系统中去。

经过编纂的知识要建立索引系统，以利于查询。对于隐性知识，经过编录后，也应建立导引系统，这在当今的信息系统中，经过研究开发，也是可以做到的。

系统中的知识有的还需要处理，如进行推理或者加以集成，可以借助一些智能技术来完成这些工作。

信息系统还应该创造一个有利于人们交流信息与观点的环境，通过信息工具，不但能使不在一个地方的人进行沟通，而且还可以保存交流的结果，甚至还能通过网络共同完成同一任务。

上面是从广义来理解知识运作与管理的技术工具这一名词的含义的。如果再细分一下，又可以分成两个层次，一是狭义的工具层次，二是方法层次。前者主要指的是信息基础设施，包括硬件和相应的服务软件，后者主要是指如何在信息基础设施上收集、处理和利用知识的方法。

信息基础设施的主要任务是：

- (1) 提供信息和知识流转的工具，以利于知识的获取与共享；
- (2) 提供知识储存的工具；
- (3) 提供知识的关联和索引工具；
- (4) 提供人员进行知识交流协作的工具。

我们先从组织的信息系统整体着眼。信息系统的组成要素不仅包括技术装备，而且包括人员及组织、工作程序和规章制度在内。这些构成了企业的信息基础设施。

本书在第2章已经介绍过信息系统与信息基础设施了，这里就不再重复。

用于知识处理与知识管理的信息系统的结构是可以分层次的。一般认为，可以分成七层，即接口层、接入与认证层、协作与过滤层、应用层、传输层、知识源集成层和知识存储层。

本书作者认为可以简化分为如图3.2所示的四层。

四个层次自下而上分别是：接口层、沟通与协作层、应用层和存储层。

主要设施有两大类，一类是包含计算机在内的信息网络，一类是基于数据库的知识库。

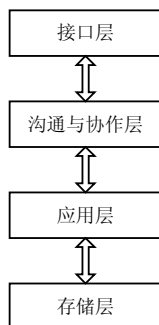


图 3.2 知识处理与知识管理信息系统的四层结构

3.6.2 知识库

如前所述,知识系统所依托的信息基础设施有两类,第一类是3.6.1节所讲的信息网络系统,第二类信息基础设施是建立在数据库技术上的知识库。为了保存信息与知识,知识系统中必须有存储设备。信息与知识存储的基础设施就是数据库及其管理系统。因为可编码的知识是以信息为媒体的,而信息的存储与流动又是以抽象的数据形式进行的。

知识库是存放知识的地方。但是这一名词却有着不同的理解。狭义的理解是属于人工智能学科的知识工程技术中,存放已经按照一定格式结构化了的知识(如专家系统中的规则)的地方。广义的理解则是包括知识获取、组织、存储、转发和维护更新所有这些功能的技术系统。显然这里存放的是显性知识。

从物理上来说,最终还是以字符或数据形式存在数据库系统之中,但对用户来说,则是按照知识表示的形式来进行交互工作的。

在知识经过收集、格式化并转换为数字形式后,可以放进知识储备库。它包含多种知识结构:

- (1) 字典,包括知识领域中名词的定义,概念和词汇;
- (2) 图像库,包括数字化的图像和视频文件;
- (3) 文本库,包括书籍、期刊、手册和说明书等;
- (4) 文档库,包括超文本;
- (5) 数据库,包括各种类型的数据库;
- (6) 实例库,包括决策与解决问题的典型事例;
- (7) 规则库,包括定义性的知识、决策与解决问题的规则;
- (8) 脚本库,包括事件、过程和典型的行为;
- (9) 对象库,包括概念、实体和对象;
- (10) 过程库,包括工作流程;
- (11) 模型库,包括因果关系模型。

为了便于知识的取用,应该以模块化、对象化的方式来进行存储。为此应该将需要存储和取用的知识分成许多基本单元。单元的划分应该考虑不同的详略程度,或者说不同的“粒度”。

粒度可以分成下列五个层次:

- (1) 知识域;
- (2) 知识区;
- (3) 知识段;
- (4) 知识元素;

(5) 知识粒子。

每一层的跨度多大，还得由知识管理与应用的范围来定。

为了便于查找，应该给不同的知识单元加上标识，其表示的属性则视知识的性质而定。

数据仓库是知识系统信息基础设施中另一种重要设备。它是计算机技术与数据应用发展到一定阶段，为满足数据的深入应用而产生的。由于企业信息化的发展，人们已经不能满足于信息系统在一般数据处理、在业务处理等方面的应用，而需要从数据中获取更有意义的信息与知识，来提高决策水平，提升企业的竞争优势。数据仓库正是为了这一目的而建立起来的一种一体化的数据存储环境，它将用于深入分析的大量数据，从传统的数据操作环境中分离出来，使分散的、不一致的用于操作的数据，转换成统一的、集成的数据。组织内不同的部门、不同的人员都可以在这样一种环境中进行分析，发现新的有用信息和知识，开阔视野，产生新的思路与方案，以提高组织的竞争能力。

数据仓库到现在还没有一个众所公认的令人满意的定义，因为数据仓库的技术与能力还在不断发展之中。当前比较一致的对数据仓库的认识是：数据仓库中的数据是面向主题的、集成的、不可更新的、随时间变化的，建立数据仓库的目的是更好地支持决策分析。

关于数据库与数据仓库及其应用，在其他课程中已经有所介绍，这里就不再赘述了。

3.7 知识经营和知识服务

要想使知识发挥作用，需要进行知识经营和知识服务。

3.7.1 知识经营

知识经济作为以知识的生产、传播和使用为基础的经济形式，它的发展一步也离不开知识经营活动。在知识经济时代，无论是生产知识的知识产业或是生产知识密集型产品的产业，都需要进行知识经营。生产出来的知识是为了使用，而要使用知识必须要有对知识进行组织、协调和在市场上交易的知识。这些就是知识经营的重要内容。所以说知识经营是知识经济发展的体现和反映。

知识经营是创造（生产）、使用、保存、提升和转让知识及智力的一种全新管理模式，是市场经济发展到一个崭新阶段的产物。知识经营把体现人的素质和智力的人力资源的主要环节，即信息技术、市场预测、经营策略、经营战略等快速而有效地统一起来，使之保持和发挥优势服务。

如果说资本经营紧紧抓住资本价值这个核心，实现了由实物管理向高层次价值管理的质的飞跃，那么知识经营则紧紧抓住知识价值这个核心，实现了由价值管理向更高层

次知识价值管理的飞跃。这次飞跃是世界进入工业社会以来一次更高层次、更高效率的飞跃。是一次充分体现经济发展知识化的飞跃，是适合知识经济发展和企业知识化推进而出现的一种新的管理战略。

知识产品的生产就是知识的创新。知识产品的生产者有不同的类型：有个体化的生产者，如个体软件开发者；有企业化的生产者，即由企业将不同的知识工作者组成不同的团队，分工合作来完成知识产品的生产，如大型新产品的开发和设计；还有国家化（甚至于国际化）的生产者，对于特大型的研究开发项目需要在全国范围内组织协作，有些关乎全人类健康的医疗卫生、环保研究项目，甚至要由许多国家联合攻关。

知识产品市场是社会分工的进一步发展，是由于知识产品成为特殊商品而产生的。

知识产品市场有其自己的特点，与物质产品市场不完全一样。它的形式多种多样，如书店、书摊是知识产品市场；科技交流贸易会、科技交流服务中心、多种方式的科技咨询活动以及各种实用技术学习班等也都是知识产品市场。

在知识产品市场上交换知识产品的，不仅有知识产品生产者的同行，还有与物质产品生产者的交换。

知识产品市场有着广泛的社会作用。它有机地联系着知识的生产和使用。它不仅可以促进知识产品的生产，促进社会的精神文明建设，而且还可以推进物质产品生产的进步，提高社会生产力，促进物质文明建设。

在知识经济时代，知识将成为竞争优势。现在一些企业希望产品更新，常常去购买一些新产品的专利许可，实际上也就是在购买知识。而有一些企业就更进一步，它们不仅是获得知识，而且是获得知识的创造者，也就是设法得到创造知识的人才或团队，甚至收购掌握某些特有知识的企业。

由于显性知识可以在市场上购买，某一个企业依仗着它暂时取得的竞争优势，其他企业也会很快地在市场上买来而取得同样优势。但是隐性知识就无法买到，所以企业关注组织知识的时候，首先要关注隐性知识（诀窍和经验等）。对于人才的评价，也不完全看显性知识，也得看隐性知识。组织知识中的隐性知识，有时就是制胜的法宝。但是隐性知识又不能直接进行交换，于是就表现为对人才的交流与争夺。

知识产品的利用是一个复杂的过程。一般分成两个步骤：一是将基础研究成果转化为应用技术，二是将应用技术转化为生产力。科技的发展使科学技术逐渐融合，以加速转化。上述过程都有如何经营管理的问题，需要建立一定的组织和运行规则。

3.7.2 知识服务

知识服务就是以知识向用户提供服务。对于知识服务现在有两种不同的理解，一种是直接将知识提供给用户，如技术或管理咨询，咨询公司或者咨询专家就是直接将知识

向用户提供。另一种则是知识密集型服务，也就是服务所用的知识密度很高甚至含有一定的创新性。其实现代化的服务都有一定的知识含量，像有一些服务如金融保险、研究开发服务和医疗保健等，都是知识高度密集型（服务人员的知识水平很高的）的服务。

咨询服务是典型的直接提供知识来进行服务的。随着市场经济的不断发展，企业在形成核心竞争力和进行管理变革方面的要求越来越高，企业急需具有战略规划经验、现场管理经验、市场营销经验的人才。有时候企业的高薪都无法聘到或者留住这些人，而社会的分工就进一步形成以知识服务为特征的咨询顾问公司。它们对社会的贡献和生存基础是为其他企业提供专业服务。这个行业的经营特征是通过对外输出知识来体现自身的价值。

咨询企业的市场就是另外一些等待发展的企业，无论是企业的创业成长，还是处于破产威胁阶段，抑或有突现出来的发展机会，企业都存在种种疑惑或顾虑，都需要社会有越来越专业的咨询顾问提供服务。

图书馆的知识服务也是典型的一类以知识内容为主要“输出”的服务。而对图书馆知识服务的研究近年来在图书馆学界正方兴未艾。以往的图书馆的服务只能说是信息服务，对所采集、提供的信息，并不要求对其所含的知识内容给以具体的分析和提炼。现在正在扩展到能向用户提供知识服务，这就需要对信息资源进行深层次的开发。对于显性知识的提供服务较为单纯，而要将图书馆的专业人员的大量隐性知识也随同显性知识一同提供服务，则是一个很有挑战性的问题。

在信息技术行业，知识服务是信息服务的进一步深化。相关工作的焦点是尝试把知识技术（本体、知识工程和数据挖掘等）与面向服务的计算技术结合起来，用信息和通信技术手段向用户提供所需的知识内容。

知识服务的模式也是多种多样的，而且随着知识服务的迅速发展，会有越来越多的新模式出现。现有的模式有下列几种：从服务过程来看，一种是全程服务模式，就像律师的服务一样，要建立全程的服务责任制，从知识的需求分析和识别、知识的获取和组织、知识的集成和应用甚至必要的创新，实现全程一体化的服务；一种是专业化的服务，只按照一定的专业要求提供对口服务。从服务人员的组成来看，一种是个人服务，由一位专家自始至终负责；一种是团队化服务，这在服务涉及的专业较广而服务人员的知识有限时，运用多个专家的力量不但能够覆盖较广的专业领域，而且还可以发挥组织知识的作用。服务可能是有偿的，如咨询公司，它们的服务是要进入知识市场的；也可能是无偿的，如政府机关提供的某些服务，图书馆的服务，以及某些社会组织帮助弱势群体提供的服务。

上面谈到的都是专门从事服务的企业或者以服务为业的专业人员。其实也有一些机构或者人员，部分承担知识服务的任务。如果从更广义的视角来看，把服务理解为某个

人员或某一组织为其他人或其他组织做事的话，在一个组织内部或个人之间，也存在知识服务。国内外都有人对于知识管理这一称呼感到不够满意，因为管理意味着控制和约束，人们本来是希望让知识更多地发挥作用，所以建议把这一管理领域称为“知识致能”或“知识促效”。其实称为广义的知识服务也许是恰当的，因为服务含有主动性，这样也许使得知识提供者更有责任感一些。

关于“知识密集型服务”的确切含义，当前人们尚未取得充分的共识。一般可以认为，知识密集型服务的关键特征在于知识是服务的重要“输入”，知识在服务提供中发挥关键性作用，但输出的知识是嵌入在其他服务活动或服务成果之中的。按照这一理解，“知识密集型服务”就涵盖了现代服务业中的很大门类，这样又与整个现代服务业联系起来了。

3.8 知识的社会与文化环境

知识系统和组织文化有着密切关系。组织中知识的共享和知识管理的成功大部分与组织文化相关，成功的知识管理必须依赖文化、管理和组织三个层面上的配合。

组织文化是组织成员在长期适应外界环境、整合内部组织的过程中形成的全体人员共同的价值观体系，是全体人员共同的精神、观念、作风、心理和习惯的总和。

组织文化是文化对组织进行渗透的结果，是文化融汇于管理之中而形成的管理与文化的共同体。文化是人类在社会实践过程中所获得的能力和创造成果。组织文化并不追求政治权力，而是反映组织这个实体的价值观、目的、要求、行为准则和习惯，不是通过行政命令能够形成的。组织文化是通过对人这一能动主体的研究，更完满、更充分发挥人的主动性和创造性而形成的。

一般来说，一个文化系统可以分为技术、制度和观念三个层面，从这三个层面来看，组织文化可以分成组织的物质文化、组织的行为文化和组织的精神文化三种类型。

1) 组织的物质文化

组织的物质文化是以物质为形态的表层组织文化，包括工作环境和生活环境，组织的器物（主要是组织的产品）反映出来的文化和组织的标识（组织的名称、组织的象征物）等。

2) 组织的行为文化

组织的行为文化是组织中的成员在工作和人际关系中产生的，表现在活动中的文化。它是以人的行为作为形态的。它包括组织的目标、组织的制度、组织的民主、组织的文化活动和组织的人际关系。

3) 组织的精神文化

组织的精神文化是在工作中形成的组织意识和文化观念，是以人的意识为形态的，

包括组织的价值观、组织的精神、组织的道德和组织的哲学。它是组织文化心理积淀的一种群体意识。

从文化的层次来看,组织的物质文化是一种表层的文化,是组织的形象。组织的行为文化是一种中层的文化,是为了统一整个组织的思想,为完成组织的目标而制定的一系列规章制度。除了成文的规章制度外,还有一些在实践中形成的不成文的规矩或者潜规则,人们按照它行事,久而久之有的也就成为成文的规章制度。组织的精神文化是深层文化,是组织中的共同信念和信仰。

在一定的组织文化环境下,组织文化对知识系统的影响,表现在下面三方面。

1) 组织文化影响员工的个人知识与公司集体知识之间的关系

组织文化界定了哪些知识属于组织,哪些知识属于员工个人;还决定了组织中什么人应该具备什么知识,谁来保存知识,谁来保护知识,谁能分享哪些知识。表面看来这是由组织的制度文化确定的,但实际上组织的精神文化在深层次起作用。如果组织领导的价值观中缺少对知识价值的正确认识,他也就不会认真地去识别本单位需要哪些组织知识,什么人具有组织需要的知识。他也不会从制度文化方面考虑怎样有意识地获取有用的知识,在组织内部交流有用的知识。

从广大员工的角度来说,如果缺少正确的知识价值观,也就不会主动去获取和认真地利用知识。在这样的组织中,知识与知识工作者得不到应有的尊重和重视,也不会形成合作的氛围,推动知识的交流与共享。

更重要的是如果缺少重视知识的文化氛围,组织知识只是少数人关心的对象,无法把广大员工的个人知识,特别是包括经验、诀窍在内的隐性知识,转化为作为集体财富的组织知识。

2) 组织文化为知识的应用与创新提供了软环境

组织文化为知识交流提供了良好的环境。这种交流可以是垂直的,也可以是水平的。

垂直交流发生在上下级之间。组织文化决定了领导的可接近程度,良好的组织文化使得领导和员工之间处于平等关系,员工乐于和领导进行知识的交流。现代化的企业中,领导不一定处处比下属高明,它需要时时向下属学习,获取知识与信息。如果领导平易近人,在组织中形成一种轻松的交流气氛,知识就很容易传播和共享,新知识也更容易形成。如果总是居高临下,使下属产生畏惧心理,对领导敬而远之,就根本谈不上知识交流了。

水平交流发生在员工之间。一个具有良好文化环境的团队,是不断在交流、讨论的过程中完成任务的。组织的行为规范与工作惯例对这种交流的频度与深度有很大影响。

所有的交流都为知识的有效应用与创新提供了条件。正因为如此,营造一个有利于交流的组织文化环境,是极为重要的。

此外，一个宽松的环境也是培育创新的条件。

3) 企业文化决定了对知识应用和创新的奖惩措施

为推动知识应用和创新，组织应该设立奖惩措施。这一方面是由制度文化决定的，另一方面更重要的是深层的精神文化在影响着奖惩制度的确立与实行。领导和员工思想深处对知识的态度决定了奖惩措施能否实行，以及实行以后效果如何。

上面所说的都是文化环境对知识应用和创新的影响。但是知识的传播和创新又会对组织文化的营造和发展起到推动作用。无论是领导还是一般员工，在组织中的知识得到传播、有效的知识得到应用、知识创新精神得到发扬之后，人们的精神面貌将会发生变化，组织文化必然也会受到影响而变得更加适应组织的发展。

不同的文化对知识的应用和创新有着不同的影响。由于历史、地理、经济、社会等原因，东方与西方的文化存在着差异。在当今全球化的形势下，东西双方的文化背景都会影响到中国的知识管理。

组织文化在知识的组织与管理过程中，有着显著影响，甚至有的国外学者认为，知识管理成败的90%取决于组织文化。



思考与讨论题

- (1) 知识管理与信息管理的差别何在？两者又有什么关系？
- (2) 除了利用现代信息技术来管理显性知识以外，你还能举出一些不用信息技术的知识管理的例子吗？
- (3) 隐性知识管理的难点何在？
- (4) 有人说现在是信息泛滥而知识贫乏，你同意吗？请阐明你的观点。
- (5) 怎样营造一个激励人们积极参加知识创新的环境？
- (6) 试举出一些可以由公众共享的知识的例子。
- (7) 你觉得当前在企业推行知识管理的障碍是什么？
- (8) 你觉得知识管理是“对知识进行管理”还是“用知识进行管理”？两者的差别在哪里？又有什么联系？



参考文献

- [1] 王众托. 知识系统工程[M]. 北京：科学出版社，2004.
- [2] 王众托. 知识管理[M]. 北京：科学出版社，2009.
- [3] 柯平. 知识管理学[M]. 北京：科学出版社，2007.

[4] 经济合作与发展组织 (OECD). 以知识为基础的经济[M]. 北京: 机械工业出版社, 1997.

[5] Polanyi M. The Tacit Dimensions [M]. New York: Doubleday Anchor, 1966.

[6] Steward T. Intellectual Capital [M]. New York: Doubleday/Currency, 1997.

[7] Liebowitz J. Knowledge Management Handbook [M]. London: CRC Press, 1999.

[8] 邱均平. 知识管理学[M]. 北京: 科学技术文献出版社, 2006.

[9] Wiig K. Knowledge Management Foundations [M]. Arlington TX USA: Schema Press, 1993.

[10] Beckman T. Knowledge Management Seminar notes [M]. ITESM, Monterrey, Mexico, 1998.

第 4 章

信息与知识资源及其采集

内容提要

著名管理学家彼得·德鲁克曾经说过：“现在真正控制资源的决定性的‘生产要素’既不是资本也不是土地或劳动力，而是知识”。知识作为获得竞争优势的核心资源，对其合理开发和利用，有助于提高企业的创新能力。本章从信息资源与知识资源的对比出发，给出知识资源的含义及特征，在此基础上，对知识资源进行了分类。介绍了企业知识资源的发现方法和采集方式。针对互联网上的知识资源，介绍了知识搜索引擎、问答系统、知识地图和 Agent 等相关采集技术。

本章重点

- 知识资源的概念及其分类
- 企业知识资源及其发现
- 知识采集方法和技术



4.1 信息资源

在介绍知识资源的概念之前，先来了解一下什么是信息资源。

4.1.1 信息资源的概念

所谓资源是指一切可被人类开发和利用的客观存在。信息资源（information resources）则是指人类在各种活动中产生的、有使用价值的信息的总和。信息资源广泛存在于经济、社会各个领域和部门，是各种事物形态、内在规律和其他事物联系等各种条件、关系的反映。

由于研究角度不同、学科领域不同，对于信息资源的解释也有所不同^[1,2]。

第2章曾经提到，信息资源有广义和狭义之分。而从不同学科领域来看，技术领域的解释是：信息资源是指以文字、图形、图像、声音、动画和视像等形式储存在一定的载体上并可供利用的信息；经济学领域的解释是：信息资源是有用的，能为使用者带来效益，但它的开发、加工和利用是有成本的。

对于信息资源的理解还可从数据、信息、知识三者之间的差别着眼。

1. 信息资源与数据资源的区别

数据是客观世界的反映，数据本身并不一定是信息，数据成为信息的前提有两个，或者是数据本身直接有用，或者数据经过加工处理后的结果就变成了信息。因此，相同的数据，不同的人、不同的处理方法可以得到不同的信息。

2. 信息资源与知识资源的区别

从信息与知识的区别来看，信息是人们感知并反映出的事物运动的状态及其变化方式，而知识则是人们对自然界和人类社会的反映和经验的总结，属于认知的范畴。因此，分别以信息和知识为起始要素的信息资源和知识资源从本质上讲分属于两个不同的范畴。但由于信息与知识之间存在着转化关系，信息资源经整理和加工有可能成为知识资源，而知识资源在被用户需求时有可能成为信息资源。这种转化关系使得现实生活中的信息资源和知识资源会出现重叠和交叉，在使用过程中较难分清哪些是信息资源，哪些是知识资源。信息资源与知识资源是两个既有联系又有区别的两个概念，虽然在一些情况下它们所指的含义是相近甚至是相同的，但是在很多情况下它们又有各自的表达领域和所指范围。

4.1.2 信息资源的特征

信息资源作为一种特殊资源，它具有与自然资源、物质资源、人力资源和财力资源等不同的特征，尤其是与物质资源相比，其独特性表现在依附性、转换性、传递性、共

享性、时效性、无限性和增值性七个方面^[2]。

(1) 依附性, 即与载体的不可分性。物质资源有着具体的形态, 信息资源则较为抽象。信息资源自身不能存在和交流, 是以一定的符号系统固化在一定的物质载体上的。信息表现的是附着于脱离开事物的载体, 如文献、通信网络和数据库等, 并且必须借助于物质载体来传播。

(2) 转换性, 即信息资源的可变换、加工和转移载体的特性。信息内容在转换载体时的不变性, 使得信息可以从一种形态转换为另一种形态, 如物质信息可以转换为语言、文字、数据和图像等形式, 也可转换为计算机语言和电信信号等。同一信息可以用多种不同的载体来承载, 不同的信息也可以用同一类载体来表现。由于信息资源的这种转换性, 人们可以通过各种印刷媒介和电子媒介等, 更有效地加工处理、开发利用信息资源。

(3) 传递性, 即易传输性。信息在时间上的传递就是存储, 在空间上的传递就是扩散。现代信息技术的飞速发展使信息传输网络变得更加高速和便捷, 使信息资源可在瞬间传递到遥远的信息消费者手中。

(4) 共享性, 即信息资源可以为一切人所享用。在物质商品的交换中, 双方的交换是对等的。而在信息的交换中, 双方的交换是不对等的, 卖方并不失去信息, 如果交换双方没有设定限制, 卖方仍然可以使用这一信息, 仍然可以与其他买方进行交换。这种交换的不对等性, 使信息资源不能像物质资源那样被占有或拥有。信息资源与信息一样, 在消费和使用中具有非排他性的特点。信息资源的建立就是为了持续不断的反复使用, 禁止或不允许他人使用信息资源不仅是很困难的, 而且也违反了信息的本质属性。信息用户的持续增加不会降低信息资源的价值。信息只有在成为产权的情况下才会具有排他性, 如商业秘密类信息资源, 享有知识产权保护的信息资源等。它们在一段特定的时间内表现为垄断和占有, 但是用发展的观点从整个流通和使用的历史来看, 它们也是要被共享的。

(5) 时效性, 即信息资源的价值对时间的灵敏度很高。信息资源在过去、现在、将来三种时态上的配置, 对其效益的发挥影响极大。一般地说, 信息的使用价值与它的传递速度成正比, 与积压时间成反比。因此, 采集和加工信息要及时和迅速, 利用信息更要及时和迅速, 才会产生信息的最佳使用价值。但也有例外, 如收藏、鉴赏和考古等种类的信息产品, 它们的使用价值会随着时间的流逝而增长。

(6) 无限性, 即就其时间的延续和信息资源的储量而言是无限的。信息资源产生并作用于人类的社会实践活动。人类的社会实践活动是一个永不停息的过程, 信息资源也总是呈现出不断丰富、不断增长的趋势。与物质资源最大的不同点是, 信息资源是可再生的, 可以被多次开发和多次利用。经过多次开发利用, 信息资源的储量不会因此而减少, 反而会不断得到扩充, 其扩充速度也会不断加快。从这个意义上说, 信息资源是取

之不尽、用之不竭的。

(7) 增值性, 即信息资源的投入不但可以使自然资源、人力资源和资本资源增值, 同时自身也具有增值的特性。社会经济发展到现阶段, 信息资源已成为一项重要的生产投入要素, 它对其他要素具有先导和决定性的影响。同时, 信息资源也使自身增值, 知识通过价值链一次又一次地被附加上去, 数据处理系统、管理信息系统演进为知识系统、专家系统, 信息产业成为低消耗和高增值的产业。

4.1.3 信息资源的分类

信息资源大致可分为知识型信息资源、经济型信息资源和社会型信息资源三类^[1]。

(1) 知识型信息资源, 其特点是提供知识获取方法的信息。它主要有科技文献信息和图书情报信息, 至于其具体内容, 则更多的是知识。比如, 图书馆有一本关于信息系统开发的书籍, 这是信息, 但是书中具体的系统开发方法、技术和工具等内容, 则是知识。

(2) 经济型信息资源, 是指对经济增长有直接作用的信息, 如市场信息和产品信息等。

(3) 社会型信息资源, 是指对社会进步、群众福利等起作用的信息。注意, 还有相当部分的在特定条件下与人类社会生活无关的信息, 不能视为信息资源, 如纯科学领域自然界活动中的大量信息, 人类只能监控到其中的极小部分。

4.2 知识资源

前面曾经提到, 信息与知识之间是可以互相转化的, 这种转化特性使得我们有时较难区分哪些是信息资源, 哪些是知识资源。下面从知识本身具有的特性出发, 来介绍知识资源的概念和构成模型。

4.2.1 知识资源的概念

知识资源 (knowledge resources) 是人类在长期实践中形成的对社会和自然的认识及判断在不断搜集、保护和追加性积累下形成的体系化总汇, 能够被人类利用到各种社会活动中作为价值创造的非物质源泉^[3]。知识资源也有狭义的和广义的两种理解。狭义的知识资源仅指知识本身, 广义的知识资源是指除知识本身以外的资源, 还包括知识产品 (如书刊、软件等)、知识工具 (如知识管理系统) 以及其他一些无形资源 (如声誉和形象)。换句话说, 广义的知识资源包括已经被占有的知识 (知识资产) 在内的所有知识资源。

为了更好地理解知识资源的内涵, 下面将其与物质资源进行对比分析。

(1) 知识作为企业中的一项重要资源, 与一般的自然资源或物质形态的资源有重大区别。首先, 作为物质形态的资源永远具有稀缺性, 即便不是绝对稀缺, 起码也是相对

稀缺。经济学所要解决的核心问题，就是以最小的投入得到最大的产出，其缘由在于资源的稀缺性。知识资源具有丰富性，人们的想象力无限、创造力无限，由此创造的知识也是无限的。

(2) 物质形态的资源具有强烈的排他性。但知识资源不排他，可以共享。

(3) 物质资源在运动、交换和变化的过程中，通常只发生价值形态的转移。比如，从实物形态变成货币形态。知识资源则不然，在运动、交换和变换过程中不仅有价值形态的转移，更重要的是产生增量，发生价值增值。英国大文豪萧伯纳讲过一句话：你有一个苹果我有一个苹果，彼此交换后，每个人手上仍然只有一个苹果。你有一个思想，我有一个思想，彼此交换后每个人都有两个思想。作为物质形态的苹果，在交换过程中一个还是一个；而作为知识形态的思想，在交换过程中从“一”增加为“二”。

(4) 物质资源投入产出的关系是清晰的、确定的和可以预见的，甚至可以用数学公式来表达。然而，知识资源的投入产出关系不清晰，具有高度的模糊性和不确定性。其情形可能是：投入巨大产出亦巨大；投入巨大但产出甚微。

(5) 物质资源的重置成本或边际成本很高，差不多与初始成本相当。然而，知识资源的重置成本或边际成本却几乎为零（尽管其初始成本往往很高）。

(6) 物质资源随着使用次数的增加会发生损耗、折旧和效用递减，直至消耗殆尽、废弃或自行衰亡，一般不会再生产，且在使用过程中或多或少会对自然环境和生态造成伤害甚至破坏。知识资源不会损耗，可以多次使用，重复交易，且知识资源随使用频率的增加其效用随之增加。知识资源不会自行消亡，可以穿越时空，长存永生；在使用时也不会伤害自然环境和生态。

(7) 对物质资源加工后形成的产品，通常只能实现某种功能或效用，具有某种使用价值；然而，知识资源则可更多地产生精神价值或观念价值，可以为使用者带来精神的愉悦和心灵的慰藉，可以提升使用者的品位和情趣。

4.2.2 知识资源的特征

与信息资源和物质资源不同，知识资源具有独特的性质，表现为：

(1) 知识资源是人的智力所创造或发现的。对于存在于个人或组织之中的知识资源，是带有个性的。

(2) 知识资源可以重复利用，有的可以经久使用，有的会慢慢衰退。知识资源的组合和加工可以成为新的知识资源。

(3) 知识资产以无形的知识形态存在，需要将其转化为能够满足人们需要的物质形态（如新产品）或者精神产品（如文艺作品）才能带来力量和财富。

(4) 知识资源的运动不像物质资源的运动那样，带有明显的数量与质量特征，而是

带无形渗透特点。

(5) 知识资源是个体知识的系统化产物,是新知识产生的源泉,具有历史性、积累性、相承性和兼容性。

(6) 知识资源更具创新价值。知识资源是各个领域人类认识和经验的总结及提炼,更是各类创新活动的原始积累。

(7) 知识资源更具有保存价值。知识资源是人类社会文化、科学和非物质文化遗产的表现形式,与信息资源相比,更值得妥善分类存储。

(8) 知识资源更具有社会指导价值。知识资源由人创造,是对社会及自然的各种活动和现象的认识,与信息资源相比,对人类活动的指导价值更大。

(9) 知识资源的价值更多地体现为人类社会创造性活动的源泉以及人类的记忆系统,而信息资源的价值更多地体现为对特定需求的满足。就此而言,知识资源在国家层面上的意义更大。

微观到企业,它的知识资源是指其拥有的可以反复利用的,建立在知识和信息技术基础上的,能给企业带来财富增长的一类资源。企业知识资源具有以下五个特点。

(1) 企业知识资源是赋予并表现企业个性的资源。不同的企业拥有不同的知识资源,不同企业的知识资源形成各自独有的文化氛围,表现出不同的企业个性和特征。

(2) 企业知识资源是市场交换性较低的资源。因为知识资源经常是某个企业所独有的,是企业的个性,所以知识资源以整体形式进入市场进行交换的可能性比较小。这包括企业的信誉、经营方式和习惯、员工的素质等。

(3) 企业知识资源是可以增殖的。在企业的运营过程中新的知识不断地生产出来,并与原有的知识资源结合,在此基础上产生更新更多的知识资源。

(4) 企业知识资源具有价值和使用价值。经过与物质结合,知识资源可以转化为物质财产,人们通过对知识资源的开发可以拥有知识资源的财产权,这部分具有财产权的知识资源就变成了企业的知识资产。

(5) 企业知识资源是可以在企业成长的同时不断地再生产出来,并与原有的知识资源重新组合,增殖出新的具有成长能力的资源。企业知识资源是活的、动态地存在。

4.2.3 知识资源模型

为深入理解知识资源的概念,下面以图形化方式,按照从核心结构到产生过程、从价值实现到配置优化的逻辑思路,对知识资源进行建模^[4]。

1. 组成要素模型

系统通常由若干要素构成,要素之间的相互作用直接影响着系统整体功能的发挥,

研究某一对象也就应该从分析其内部构造入手,展开对整体运行规律的探寻。知识资源除了知识自身,还包括知识人(生产者、加工者、传递者和利用者等),知识工具(方法、设施和网络等),以及知识活动(生产、流通、分配、服务、开发和交流等),整个体系以科学知识为基础,以知识创新为宗旨,以价值创造为导向,各组成要素的结构关系——知识资源钟如图4.1所示。

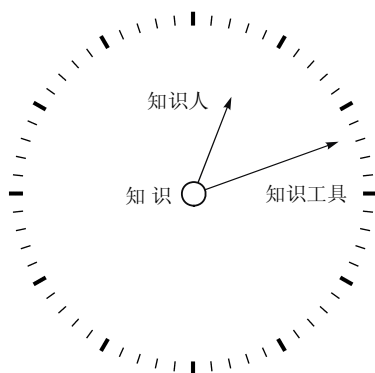


图 4.1 知识资源钟（一种隐喻结构模型）

作为一种隐喻模型,图4.1中时钟的各部件分别有其代表含义:钟面代表信息,刻度代表价值指标,中轴代表知识,时针代表知识人,分针代表知识工具。这个象征性模型体现了知识资源各组成要素及其相互之间的协同关系和影响作用。整个时钟从本体论的角度模拟知识资源系统的构成原理,它具体包含了以下内容。

(1) 如同时针和分针以中轴为圆心旋转一样,知识人和知识工具围绕知识开展活动,以其为行为主轴,以此确立知识资源的核心要素。

(2) 指针的运行脱离不了盘面的支撑,知识活动建立在信息的基础之上,这是知识资源最直接的外部交换系统。

(3) 时针与分针的相互影响体现了“工具”与“人”的辩证关系。一方面,如果缺乏知识工具,知识人所能产生的知识价值将不够高效和精准;另一方面,知识工具以知识人为主导,为知识人服务,离开知识人的知识工具将毫无意义。

(4) 时钟的运转体现了知识资源的社会价值通过知识活动得以实现,并且是一个循环反复、螺旋上升的过程。

2. 资源形成模型

知识资源不仅仅是单纯的知识本身,进入人类知识体系中的内容必须经过一系列的组织、管理、控制和开发的过程才能为生产力的发展提供能量基础,为社会的进步提供财富来源。这一过程离不开从理论研究到技术革新,再到实践应用的递进发展,离不开学术研究对资源创造的积极影响。知识资源形成的影响因素较为复杂,图4.2从科研活

动的视角将主要过程进行了概括。

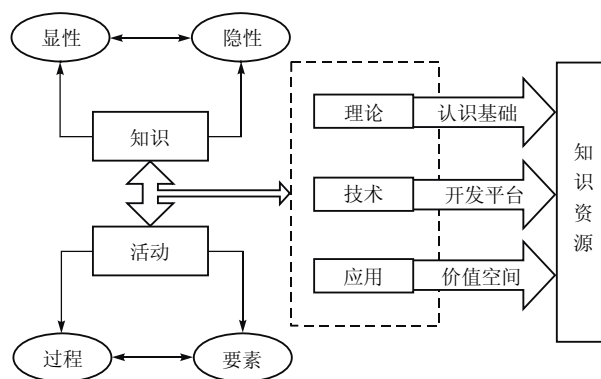


图 4.2 知识资源形成模型

根据图 4.2 的显示，显性知识与隐性知识交互转化，不断地丰富着人类知识的总量，与知识相关的各种管理控制活动则促使静态的已有知识通过吸收和创新的过程，利用恰当的技术和工具，应用到生产实践之中。这些阶段和环节最终形成的知识资源，为社会效益的创造提供了认识基础、开发平台和价值空间。

这些来自于不同层面的知识资源具备不同的特性，如果深入分析，各自的形成原理不尽相同，在资源总库中所承担的功能和效用也各具特点。

3. 价值实现模型

知识资源的构成原理与形成过程固然重要，但其社会价值的实现才是存在的依据和根本。研究知识资源的基本功能首先要强调的就是资源效益的产生原理及相关问题，这反映了从本体论上升到认识论的研究过程。

(1) 微观层面的价值实现。知识的生产管理、组织加工以及开发利用和功能发挥历来受到人们的重视。从知识的形成到最终成为社会生产力的一部分，可以从微观流程的层面来理解其主要功能的发挥，这是一个由自然事物属性向社会记忆系统转换，再向智力资源体系的演化进程，图 4.3 以金字塔的形状构建阐释了其中的各个发展环节。

事实是人类对客观存在的思想映射，是人脑对事物属性做出反应的结果。数据是客观事实编码化、序列化的结果，是形成智力资产的具有初步价值的基本要素。信息是记录或承载于媒介之上的数据，是形成人类知识的直接来源。知识是经过加工、提取和评价的信息，是形成资源的核心要素。决策是利用知识解决问题的智能活动，是价值实现的必备前提。竞争优势是通过决策确立的保持领先的创新能力，是组织或个人利用知识资源的效益体现。

由金字塔的塔基向顶端的递进代表了知识资源的价值实现过程。客观属性内化是事物本质属性被人类的思维活动条理化的过程，通过已有的知识框架对感官受到的刺激进

行编码和整序，形成底层的智力资产。媒介载体物化是将原始数据信息化的过程，它将包括人工符号、比特字节在内的各种内容特征信号按一定标准进行处理，形成便于识别的信息。加工提取集成是由信息向知识转换的过程，它以真实客观为标准，促使信息中的价值成分充分体现，形成能够指导社会实践的知识。问题导向智能是利用知识进行参考决策的过程，以特定问题的解决为目标，运用人类智力，开发知识价值，寻求解决方案。协同行动管理是通过实施决策方案，建立竞争优势的过程，体现以知识资源为基础的独特优势，完成知识资源的价值实现。

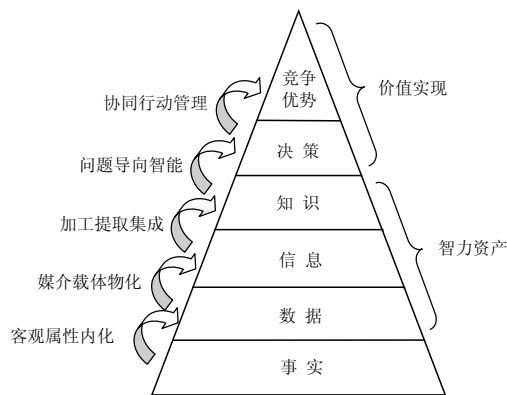


图 4.3 知识资源演化模型

（2）宏观层面的价值实现。无论信息、知识，还是资源，一个共同的现实存在的前提条件就是由诸多相关因素构成的生态环境，与社会生产发生直接联系的知识资源尤其需要强调适应外部环境的演化变迁，强调影响因素对知识活动的作用反馈，强调知识资源与其他生态系统在能量流动和物质循环的过程中达到持续稳定的平衡状态。从宏观生态的视角观察知识资源的价值实现，大致的宏观模型如图 4.4 所示。

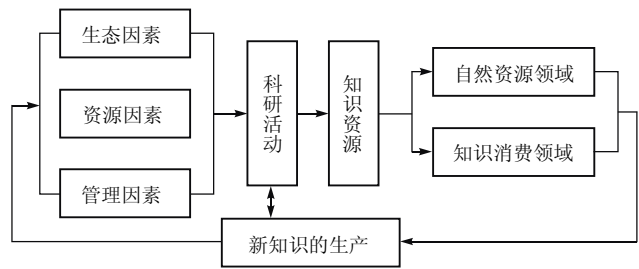


图 4.4 知识资源价值实现宏观模型

资源价值的实现包括两大领域：自然资源领域和知识消费领域。在整个生态环境中，人类知识对社会进步和经济发展的推动作用首先体现在促使自然物质的效用开发和功能转换，以此探求新型的非稀缺的自然资源；其次就是满足人类日益高涨的知识需求，这必然导致知识产品的大量涌现和传统商品中知识含量的迅猛增长。这些领域的演化和拓

展均毫不例外地刺激着新知识的诞生，推动着人类知识的积累升华，并促使其再次进入由管理、技术服务等社会因素构成的生态环境，经过科学活动的再加工，形成新一轮的知识资源，在动态转换中寻求着最优的共生关系，在持续的价值实现进程中提升着人类智力的总体水平。

4. 资源配置模型

在研究自然资源功能发挥与性价比优化的过程中，人们发现开发利用与实际需求之间并非均衡对等，经常性的失调现象使得资源配置的问题日显突出，知识资源的研究与利用同样需要处理好开发与需求的相互关系，两者由低向高的递进可以形成如图 4.5 所示的知识资源矩阵模型。

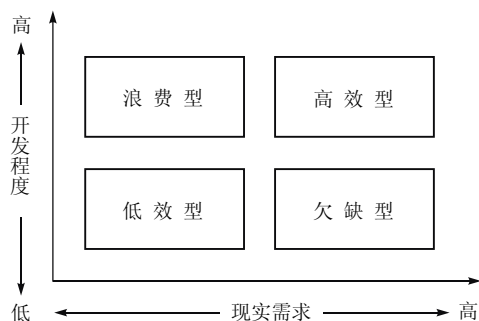


图 4.5 知识资源矩阵模型

矩阵中不同的组合形式代表了四种知识资源的配置状态，一是低效型，属于开发程度和现实需求都较低的类型，常见于物质资源和文明程度均不发达的地区或时期，是知识资源价值体现的原始阶段。二是欠缺型，属于开发程度低而现实需求高的类型，常见于受限于方法技术的发展阶段，是对知识资源需求超前，但开发手段相对落后的意识先导状态。三是浪费型，属于开发程度高而现实需求低的类型，往往由过分强调和依赖技术所致，是产品超越需求的技术先导状态，当前电子产品的频繁升级、功能叠加便是典型的例子。四是高效型，属于开发程度和现实需求都较高的类型，知识资源的价值得以充分而恰当地体现，是最佳的资源配置状态。

4.3 知识资源的分类

一般来说，知识资源按照不同的标准可以进行不同的划分。

(1) 按知识的形态划分，可分为隐性知识 (tacit knowledge) 资源和显性知识 (explicit knowledge) 资源。隐性知识是高度个人化的知识，有其自身的特殊含义，因此很难规范化，也不容易传递给他人。隐性知识是一种主观的、基于长期经验积累的知识，它不能

用几个词、几句话、几组数据或公式来表达,它的内容都有十分特殊的含义。这类知识主要包括信仰、隐喻、直觉、思维模式和所谓的“诀窍”(如企业的管理理念、一些应急处置方案)。而显性知识是指可以通过规范化和系统化的语言、文本方式进行记载和传播的知识,可以采用现代技术手段进行收集和处理。典型的显性知识主要有专利、文件、规章制度、设计图、论文和报告等。企业的显性知识资源主要包括:人力资源,包括企业的专家、技术人才和高质量的劳动者;基础设施资源,包括企业文化、金融结构、数据库和通信系统等;市场资源,包括企业的品牌、信誉、客户、销售渠道和许可证协议等;知识产权资源,包括企业具有所有权的专利、版权、商标、商业秘密和专业技术等。

(2) 按共享程度划分,可分为共享知识资源和非共享知识资源;公共物品性质知识资源(涉及国家安全和国家经济命脉的知识资源,如政府知识资源,包括图书、情报、档案、出版和发行等系列的知识资源等)和非公共物品性质知识资源(如知识产品与服务),无偿利用的知识资源和有偿利用的知识资源;公共知识资源(公益性知识资源)和非公共知识资源(或私有知识资源)等。

(3) 按学科领域划分,可分为科学知识资源、技术知识资源和信息知识资源。

(4) 按组织边界划分,可分为内部知识资源和外部知识资源。

(5) 按载体系统划分,可分为人、符号编码系统(包括文字、语言、图形、音频、视频和符号)和物质与非物质的产品系统(包括各种包含着知识和技巧的物质和非物质产品,如瓷器、计算机软件、工艺品、各种非物质文化遗产等),相应地,其知识资源也可分为三类。

(6) 按效用划分,可分为记忆性效用、创新性效用和传承性效用。具有记忆性效用的知识资源主要采取存储的战略措施,具有创新性效用的知识资源主要保证其利用与传播的有效性,而具有继承性效用的知识资源主要实现其保存活性的存储,即技艺和知识的可继承。

(7) 按内容来划分,企业知识资源可以分为:

- ① 信誉资源,包括商标、品牌、企业文化和信誉等无形资产。
- ② 技术资源,包括产品的研制、工艺、专利、版权、技术诀窍和商业秘密等。
- ③ 信息资源,通过信息网络可以收集到的与企业生产经营有关的各种信息。
- ④ 客户资源,是指企业拥有的客户关系资源。
- ⑤ 人力资源,企业中所有那些体现在企业员工身上的才能,包括企业员工的专业技能、创造力、解决问题的能力和管理者的管理能力。在某些情况下,甚至还包括企业员工的心理能力,因为企业员工的心理素质在很大程度上将影响其才能的发挥。
- ⑥ 基础结构资源,企业的基础结构主要指企业的管理哲学、企业文化和企业内部的基本管理制度以及企业与外界力量所形成的各种协议和制度安排。

4.4 知识资源的发现与采集

由于知识资源具有非物质特性，因此对它的发现和采集十分困难。

4.4.1 知识资源的发现

知识资源的发现是一个比较复杂的过程，属于高度智能化的范畴。对于显性知识资源，其发现过程是识别储存在数据库中有效的、新颖的、具有潜在价值和最终可理解的模式的过程，而数据挖掘则是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的过程。知识发现系统基于用户知识发现任务驱动，依赖领域内的背景知识，尽可能自动地、智能地、有效地挖掘数据内部未知的、潜在的模式以及趋势，使人与计算机达到理想组合。知识资源的发现是利用知识发现引擎，依托知识库，选用高效的算法，对数据进行挖掘和选择。目前已有许多较为成熟的算法，如分类、聚类、关联规则和异常点检测等。知识一旦被挖掘出来，就要对发现的知识进行评价，筛选出有意义的知识。

对于隐性知识资源的发现，是一件十分困难的事情，因为隐性知识大多存在于人们的头脑中，不能通过常规手段发现，只能通过模仿、交谈和实践等方式发现和获取。

除了上述两种知识资源的发现外，还存在第三种知识资源，即中间态知识（intermediate knowledge）^[6]。由于中间态知识既不属于显性知识，也不属于隐性知识，是最难被发现的一类知识。例如，在隐性知识转变为显性知识的过程中，从微观上分析，实际上存在两种可能。第一种情况是隐性知识不经中间态，直接发生跃迁转变为显性知识；第二个转变步骤是经过一个中间态过程，这个过程可能非常短暂，也可能经历时间较长。转换过程中存在一个临界点，处于临界点的知识属于新的显性知识的萌芽，一旦被激活，则越过临界点顺利转换为显性知识；如果未能被激活，则隐性知识不能转换为显性知识。知识转换的中间态模型如图 4.6 所示。

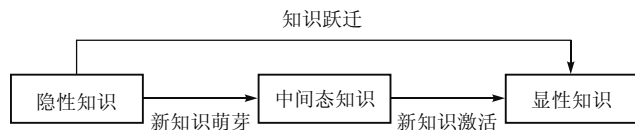


图 4.6 知识转换的中间态模型

研究知识转换的中间态，具有一定的理论和实际意义。在组织的知识管理过程中，还没有经过验证的但是已经公开提出来的假想、设想、理论、模型和流程等就不能算做组织的知识，当然也不是显性或隐性知识。它们就属于所谓的中间态知识。这种东西对

于组织的知识管理来说,是非常有意义的。因为它们——新知识的萌芽进行管理(包括提供促进这种东西产生的手段和激励机制等),显然有助于新知识的最终形成以及知识管理目标的实现。在知识管理的背景下,区别组织知识和个人知识是很有好处的。在组织中,个人隐性知识转换成显性知识的过程中,其转换的中间态显而易见是存在的,它是在组织的学习、实践过程中被激活转换的。它是一个操作性的、交流过程中的东西。中间态知识是十分重要的知识资源,但同时也是最难发现和最难控制的一类资源。

知识资源的发现是一个动态过程,人们在利用知识过程中,不断地在创造新知识、接受新知识,知识会不断自我孵化和增长。所以知识资源是一个活的体系,需要不断地发现新产生的知识。

4.4.2 知识资源的采集方式

知识资源不像其他资源那样等着被开发利用,它会能动地刺激使用者去扩张,去发现和利用那些未知的、未被利用的资源。知识资源的采集是知识利用的出发点。

企业知识资源的采集,按来源的范围大致可分为内部获取和外部获取两种。内部获取是指从企业内部的员工、技术资料、各种数据库、业务流程、日常工作历程、制度及组织文化等产生知识或蕴涵知识的地方获取知识的方式。企业内部知识资源采集的特点是时间周期长,成本不确定,形成的知识容易快速形成竞争优势,竞争者不易模仿。外部获取是指从企业以外的组织或人员获取知识的方式。外部知识资源采集一般应用于科技尚欠发达的国家和地区^[7]。

常用的采集方式有以下五种。

(1) 通过员工培训、技术设备引进和技术开发项目等方式从企业内部获得。

员工培训有很多种方法,如送人去外面培训,聘请外面的人来单位内部进行培训和教育,建立定期的知识传授与交流制度等。

(2) 通过租赁、合资、许可贸易、完全购买、引进人才和知识联盟等方式从外部获得。

租赁的方式有很多,如聘请顾问与咨询专家;从顾客、供应商、研究单位或专职机构获取帮助;与其他组织签订合同等。这对于只是一时需要的知识来说,是非常合算的。

(3) 利用文本资料,如通过各类书籍、专业期刊和大众媒体等渠道采集知识。

(4) 利用互联网采集知识,互联网是人类有史以来最大和最容易使用的百科全书,通过这种方式采集知识是非常经济和快捷的。

(5) 利用数据挖掘技术采集知识。

对于一些基础理论、涉及社会公共利益的知识,它们是人类共同财富,属于社会公共物品,人们可以不用花钱(或者花费极少的载体成本)就能够得到。

对于显性知识,我们知道互联网是一个特大的知识库,可以从网上获得国内外有关

的信息和知识。其中绝大部分是可以免费下载的，可以通过网络，使用搜索引擎和浏览器去查找有关的知识，也可以先获取有关数据（这可以通过组织中已有的信息系统中的数据库或数据仓库得到），再通过数据挖掘、知识发现等技术获得相应的知识。

对于隐性知识，由于它们难以用语言文字表述，所以很难通过搜索找到它在什么地方。因此，有效的办法是能找到谁拥有这类知识。这样一来，对知识的搜索变成了对掌握知识的人的查询。现在有一些编制知识地图的方法，用来帮助人们找到掌握相应隐性知识的人，这实际上就是建立专家网络图。

对于网络知识资源的采集，“威客(Witkey)”是一种较新的方式。威客的英文 Witkey 由 wit（智慧）和 key（钥匙）两个单词组成，也是 The key of wisdom 的缩写，是指那些通过互联网把自己的智慧、知识、能力和经验转换成实际收益的人，他们在互联网上通过解决科学、技术、工作、生活和学习中的问题，从而让知识、智慧、经验和技能体现出经济价值。利用威客，人们可以通过有偿的方式获得自己所需的知识。

4.5 知识采集技术

知识资源尤其是网络知识资源多采用浏览器、搜索引擎、智能代理和知识地图等信息技术进行采集。通过搜索引擎的特定功能或与问答系统的交互，用户可获得所需的知识资源，同时可根据链接功能将采集到的资源添加到知识地图中，帮助用户了解知识资源的分布情况。采集到的知识资源，还需对其进行真伪判断、价值评估，再按照实际情况进行加工整理，最后存放在知识库中。下面主要针对采集阶段来介绍相关技术。

4.5.1 知识搜索引擎

随着网络信息爆炸式的增长，传统的第一代和第二代搜索引擎提供给用户的查询结果往往数以万计，给用户浏览和查找有用信息带来很大的困难，因此用户希望能够拥有一种新的搜索引擎来帮助他们准确地获取所需信息，甚至帮助他们发现未知的信息或知识。不仅如此，当用户提出一个问题，这种搜索引擎还能给出最准确的答案，它就是知识搜索引擎。

知识搜索引擎是将各方面的知识资源整合在一起，不仅可以利用先进的自然语言智能查询技术，通过对输入的简单疑问词句的分析直接搜索得出答案，而且在没有满意搜索结果的情况下，用户还可以创建问题，等待他人回答并对答案进行评价。知识搜索引擎为广大网络用户创建了一个强大的信息交流平台，通过提出问题、回答问题和搜索问题来解决用户的知识需求。利用知识搜索引擎获取知识的这一过程，使得大量的隐性知识显性化，并且实现了知识的挖掘、传播、利用、共享和评价。

下面分别从搜索方式、提问方式、搜索结果、结果排序、相关协议、知识服务六个方面,介绍知识搜索引擎的特征。

1. 搜索方式

知识搜索引擎一般提供基本搜索和类目浏览搜索两种搜索方式。基本搜索即在问题搜索框中直接输入简单的词句进行搜索,而类目浏览搜索是通过类目设置逐层点击进行搜索。有的知识搜索引擎如新浪“知识人”和雅虎“知识堂”等还提供高级搜索。高级搜索可限定用户所搜索问题的类目,选择在具体的类目中进行搜索,或者限定以下内容:① 搜索范围,如限定在提问题目中搜索。② 问题分类,可以选择在全部或具体的某一个一级类目中进行搜索。③ 问题的状态,选择在全部问题、提问中、投票中或已解决的问题中进行搜索。④ 发问时间,如选择在7天以内发问的问题中进行搜索。⑤ 正面评价。有的知识搜索引擎,如雅虎“知识堂”,还提供二次搜索,可以在初次搜索结果的基础上进一步进行搜索,这样可以使用户搜索得到的结果更加精准和具有专指性。

2. 提问方式

如果通过搜索答案没有得到满意的结果,用户可通过提问的方式来获取问题的答案。一般有过客提问和注册提问两种方式。过客提问不需要注册就可以进行提问,但是只提供简单的提问设置。注册提问会提供更详细的提问设置,这样得到的答案会更确切。注册提问后,用户必须按照相关要求对自己所提出的问题和所得到的答案进行处理。提问规则如提问有一定的期限、设置提问悬赏、没有满意答案的情况下如何处理等内容事先预置于搜索引擎中,这些相关规定可以保证问题的有效性,而且可以避免问题堆积。

3. 搜索结果

利用类目浏览搜索得到的搜索结果主要包括全部问题、已解决问题、待解决问题、投票中问题、高分问题、零回答问题、高悬赏问题、推荐问题、提问中问题和知识贡献等形式;基本搜索得到的搜索结果主要包括已解决问题、待解决问题、提问中问题、投票中问题和知识贡献等形式。利用类目浏览搜索得到的搜索结果形式要比基本搜索得到的搜索结果形式多,因为利用类目浏览搜索的针对性没有基本搜索那么强,提供多种形式的搜索结果可以全面了解相关内容。这里值得一提的是高分悬赏问题,通过高分悬赏,可以吸引用户来回答问题,使得问题的答案更全面和丰富。

4. 结果排序

搜索结果一般按照相关性和发问时间进行排序,问题处于不同的状态,还有不同的排序方式。相关性,指依关键词位置、状态和时间,做权数加总的智能型排序;发问时间,指依问题的发问时间做排序(按照时间由近到远,或由远到近进行排序);赠分数(“提问中”状态),指依问题赠送的分数做排序(由多到少或由少到多);投票数(“投票中”

状态),指依问题中各答案得票总数进行排序(由多到少或由少到多);正面评价(“已解决”状态),指依最佳答案获得的正面评价的百分比进行排序(由多到少或由少到多)。

5. 相关协议

知识搜索引擎为网络用户提供了一个互助问答的平台,网络用户共同参与知识的分享和交流。为了维护网络用户的利益,保证答题的良好质量,需制定相关规定和删除标准,如百度“知道”有《知道协议》、新浪“知识人”有《删除标准》,雅虎“知识堂”有《雅虎知识堂内容删除标准》。凡是违反相关协议或是符合删除标准的提问和回答都会被强制删除,以保证交流平台的高质量、规范化和友好化。

6. 知识服务

知识搜索引擎的未来发展方向是给用户提供更多元化服务,如提供收藏夹,协助用户进行知识积累和管理;提供RSS订阅服务,使用户不断获取自己所需要的某一方面的知识;提供开放词典,使用户可以自己创建词条和编辑字典,共同建立一个知识库,供大家分享。

知识搜索引擎的优点包括如下四个方面。

(1) 搜索结果的准确性。由于采取了知识库为基础的语义分析,在进行检索过程中,采用的不是关键词全文检索,而是基于概念的检索。比如,当你输入“北京天气怎么样?”时,传统搜索引擎返回的结果连小说都检索出来,因为小说内容包括“北京天气怎么样?”这句话。而智能搜索引擎,由于采取语义分析的方法,分析出北京天气这个概念,于是直接给出北京的天气情况预报。

(2) 搜索结果的范围定位准确。由于采用知识(概念)检索技术,明确和缩小了搜索范围,减少了对无用信息范围的检索。比如,上面的例子,智能搜索引擎只在天气这个范围进行检索,从而提高了检索效率,减少了无用信息。智能搜索引擎是以搜索结果准确、范围小为特点的。

(3) 搜索结果的综合性。由于采用了知识库,搜索引擎将给用户提供更全面、更综合和更合理的知识框架,在这里,知识检索只是知识服务的一部分。比如,当你输入“在北京怎么找工作?”时,给出的答案不仅仅是给出“北京地区的招聘信息”,而且还给出“北京地区的人才政策”和“求职技巧”等信息。

(4) 搜索结果的智能性。由于有综合知识库作为背景,知识检索和导航服务变得更加智能。知识库中的知识有助于解决“表达差异”问题。例如,只要定义“计算机”、“电子计算机”、“电脑”是同义关系就可以消除用户由于使用不同的词表达同一概念而带来的检索困难。另一方面,知识库可以对用户的查询进行相关性联想,提供引导用户进行下一步查询的线索。这样一步一步地在与用户的交互过程中诱导用户“表达”出他真正想找的东西,从而实现了对查询的智能导航。这种逐步求精的策略解决了信息检索“忠实表达”的难题。

4.5.2 问答系统

问答系统（QAS, Question Answering System）也是一种知识搜索引擎。它就像一个知识渊博的专家，能用准确、简洁的自然语言回答用户用自然语言提出的问题。由于采用了语义理解技术，使得问答系统比传统的搜索引擎更方便、快捷和高效。

根据答案获取方法的不同，问答系统可分为自动问答系统和交互式问答系统两种。自动问答系统主要利用基于语义匹配的方法获取答案，而交互式问答系统则采用基于问题推荐的方法获取答案。

自动问答系统一般由问题处理模块、段落抽取/索引模块和答案处理模块三部分组成，如图 4.7 所示。

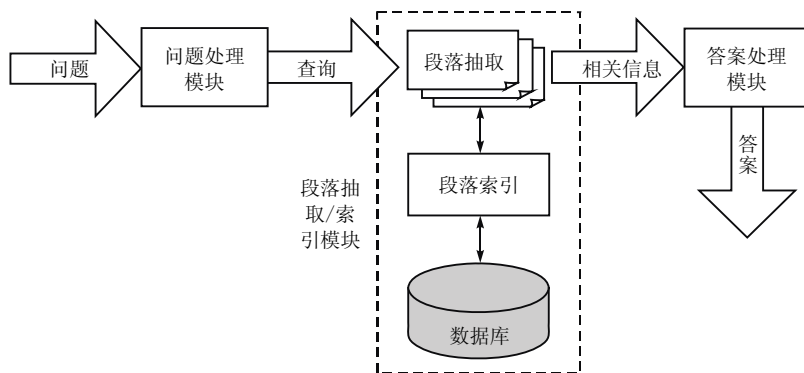


图 4.7 自动问答系统结构图

问题处理模块通过文本处理技术，分析用户提出的问题，以抽取出用户的需求并利用某种表达方式将这种需求信息表达为查询，作为段落抽取/索引模块的输入。段落索引模块负责从数据库中搜集并管理所有可利用的文档，这些文档可以从互联网中获得，也可以从本地的数据源中获得。段落抽取模块则将索引得到的文档进行分段处理；然后依据这一模块的输入，从中抽取出所有与查询相关的段落；最后对所有得到的段落进行评估，并将最相关的若干段落作为答案处理模块的输入。答案处理模块则是通过信息抽取技术从这些相关段落中抽取出若干候选答案，通过评估排序获得针对该问题较为可信的答案。

除了上述三个模块之外，有的自动问答系统还包括一个常问问题（FAQ）库，把用户经常问的问题及其答案保存起来。有了 FAQ 库之后，对用户问的问题先在 FAQ 库中搜索，看看有没有相同的问题，如果有，就可以直接把 FAQ 库中这个问题的答案返回。这样，对于用户常问的问题，自动问答系统就可以很快给出答案，而不需要经过复杂的处理，而且还能保证答案的正确。

源的总目录及各知识点间的关联；二是人员专家网络，即对企业员工的知识技能及相关领域专家的描述。利用知识地图可以清楚地知道企业内部、外部相关知识资源的分布及知识节点间的相互关联，方便对企业已积累的显性知识资源的收集与利用。至于隐性知识的采集，可以通过知识地图中的专家网络即知识与人、人与人之间的联系来获得。

知识地图的构建方法有很多种^[13]，归纳起来大致包括以下七个步骤：

- (1) 辨识知识。识别知识密集型流程，确定知识密集型区域所涉及的知识资源。
- (2) 提取知识。对选择的知识资源进行审核，然后提取知识并确定位置。
- (3) 组织知识。对提取出的知识资源进行编码、分类、标引和编辑等处理，用元知识描述知识资源，然后封装存储。
- (4) 关联知识。在知识分类的基础上，确认识知识关联。
- (5) 绘制地图。建立知识之间的关联，即形成知识网络后，便可绘制初步的知识地图。
- (6) 评价地图。验证及评价地图的有效性。
- (7) 完善地图。根据评价结果，不断修正知识地图中的内容，并在反复测评中进行完善。

4.5.4 Agent 技术

在知识采集过程中引入 Agent 技术，可帮助用户更广泛地获取知识资源。

Agent 是人工智能领域发展起来的一个概念，它是指具有感知能力、问题求解能力和与外界进行通信能力的一个实体。Agent 又称智能体，它能在用户没有明确具体要求的情况下，根据用户需要，代替用户进行各种复杂的工作，如信息查询、筛选及管理，并能推测用户的意图，自主制定、调整和执行工作计划。此外，Agent 还能从经验中不断学习。例如，Agent 代替用户在网上购物时，可以根据用户的爱好或消费水平，挑选用户喜爱的商品。通过分析用户对所选购商品的态度（如高兴、认可和不喜欢），可以进一步掌握用户购物的意向。

Agent 技术最基本的特征是：具有解决问题所需的丰富知识、策略和相关数据，能够进行相关的推理或智能计算。它利用储存在知识库里的信息执行任务，使用自动获得的领域模型（如 Web 知识、信息处理、与用户兴趣相关的信息资源、领域组织结构）、用户模型（如用户背景、兴趣、行为和风格）知识进行信息搜集、索引和过滤（包括兴趣过滤和不良信息过滤），并自动地将用户感兴趣的、对用户有用的知识提交给用户。

基于 Agent 的知识搜索系统一般包括领域模型库和用户模型库两个知识库和用户界面 Agent、学习 Agent、分析过滤 Agent、检索 Agent、选择推荐 Agent 五类 Agent 及其他辅助部分。

关于 Agent 在知识管理系统中的其他应用，将在 9.7 节中介绍。



思考与讨论题

- (1) 简述自然资源、信息资源与知识资源的联系与区别。
- (2) 知识资源如何分类? 它的特征是什么?
- (3) 企业知识资源如何界定? 采集途径有哪些?
- (4) 简述知识搜索引擎原理。对国内常用的知识搜索引擎进行比较分析。
- (5) 威客模式获取知识的优势和劣势是什么?



参考文献

- [1] 俞立平. 信息资源与自然资源内在关系及替代研究[J]. 情报资料工作, 2006, 2: 25-27.
- [2] 杨廷廷, 许宏翠. 试论信息资源的含义与特征[J]. 情报资料工作, 1998, 1: 25-27.
- [3] 王平. 论知识资源: 概念辨析及其操作化[J]. 图书情报知识, 2009, 130: 105-108.
- [4] 赵益民. 基于知识学的知识资源模型研究[J]. 图书情报知识, 2009, 127: 50-53.
- [5] 陈洪澜. 知识分类与知识资源认识论[M]. 北京: 人民出版社, 2008.
- [6] 任皓, 苏新宁, 孔敏, 等. 论企业知识资源的组织[J]. 情报学报, 2003, 22(2): 211-216.
- [7] 黄福玉, 冯玉强, 卢鹏宇. 基于资源理论观点的企业知识资源获取方法研究[J]. 2006, 情报杂志, 9: 37-39.
- [8] 王春晓. 知识搜索引擎比较分析[J]. 图书馆学研究, 2009, 1: 47-49.
- [9] 百度知道[EB/OL]. <http://zhidao.baidu.com>.
- [10] 雅虎知识堂[EB/OL]. <http://ks.cn.yahoo.com>.
- [11] Iask 爱问知识人[EB/OL]. <http://iask.sina.com.cn>.
- [12] Liu W. BuyAns—An Incentive & Collaborative Platform for Knowledge Acquisition[C]. 2006, Proceedings of 2nd International Conference on Semantics, Knowledge and Grid, GuiLin, China.
- [13] 郭佳慧. 国内外知识地图研究比较分析[J]. 2009, 29(5): 213-217.

第 5 章

知识表示

内容提要

知识表示就是对知识的一种描述，或者说是知识的一组约定，一种用于描述知识的形式和符号。对知识的表示过程就是把知识编码成某种形式和符号的过程。知识可以分为确定性和不确定性两类，相应的知识表示方法也可以分为确定性知识表示方法和不确定性知识表示方法。知识表示的方法是多种多样的，确定性知识表示方法主要有逻辑表示法、关系表示法、产生式规则表示法、框架表示法、语义网络表示法、面向对象表示法和本体表示法等；不确定性知识表示方法主要有基于概率、模糊集和粗糙集等不确定数学理论的知识表示方法。随着全球信息化进程的加快，非结构化信息处理已逐渐成为当今信息处理的主题，非结构化信息的知识表示方法是与之相关的核心科学问题之一，近年来也取得了重要的进展。本章内容是为那些未学过知识工程、人工智能的学习者提供的初步知识。

本章重点

- 确定性知识表示方法
- 不确定性知识表示方法
- 非结构化信息的知识表示方法



5.1 知识表示简介

知识表示是为描述知识而作的一组约定，是知识的形式化过程和符号化过程。从历史上来看，人们最初用来描述客观世界和人与人之间互相交往用的是自然语言。但是随着科学技术的发展和认识的深入，人们又利用符号语言和图形语言来对客观事物的本质属性以及事物间的关系加以描述。由于任何事物及其运动的属性都是质和量的统一，所以知识的描述既有定性的一面，又有定量的一面。知识的定量描述可以借助于符号语言和图形语言，而知识的定性描述主要还得借助于自然语言。

研究知识表示，就是设法找到可行的、有效的和通用的方式，既便于口头交流，又能够用工具（如在计算机上）表达。使用自然语言进行口头交流已经是日常生活中不可缺少的部分，但由于现在的交流很大部分是通过信息系统，还需要进行必要的处理，因此探讨知识形式化和符号化的方式和方法就显得特别重要^[1,2]。

20 世纪 70 年代，由于研究人工智能和建造专家系统的需要而出现了知识工程学科^[3]。所谓知识工程，是指以知识本身为处理对象，研究如何运用人工智能理论和软件工程技术，设计、构造和维护知识系统的一门学科，与此有关的理论技术、方法和工具都是知识工程的研究内容。知识表示是知识工程中最重要的问题之一。

目前，人们从不同的应用领域出发，提出了多种不同的知识表示方法。每种表示方法都有自己的特点和局限性。对同一种知识可以用不同的方法进行表示，但表示的方法不同则表示的效果也不同，因此，在选择知识表示的方法时应考虑以下几个因素：① 能否充分表示相关领域的知识；② 是否有利于对知识的利用；③ 是否便于知识的组织和管理；④ 是否便于理解和实现。

5.2 确定性与不确定性知识

知识可以分为确定性与不确定性两类。

5.2.1 确定性知识

以经典物理学理论体系为代表的确定性科学，创造了精确描述客观世界的方法。确定性科学将整个宇宙看作是机械的动力学系统，处于确定有序的运动之中，只要知道系统的初始状态、边界条件和动力学方程就可以决定系统未来的一切。从牛顿到拉普拉斯，再到爱因斯坦，他们创建的理论体系所描绘的都是一幅幅完全确定的科学图景^[4]。

在确定性科学中，人们主要利用以数学方程式为主要表现形式的符号语言来描述客观世界。由于代数方程、逻辑方程、微分方程和差分方程的产生和发展，一些客观事物中各因素的相互联系与制约、在空间的分布和随时间变化的情况，均可用上述方程来描

述。通过这些方程,可以从已知的状态或过程去推算未知的状态或过程。除了符号语言,图形语言也用得很广泛,如交通管制的路标、计算机程序框图和电子线路图等。图论是专门以图为研究对象的一个组合数学分支。图论中的图是由若干给定的点及连接两点的线所构成的图形,这种图形通常用来描述某些事物之间的某种特定关系,用点代表事物,用连接两点的线表示相应两个事物间具有这种关系。

用精确定义的概念和严格证明的定理描述现实的数量关系和空间形式,用精确控制的实验方法和精确的测量计算探索客观世界的规律,建立严密的理论体系,这是确定性科学的特点。由于确定性科学的深远影响,以至于在相当长的一段时间内,限制了人们认识宇宙的方式和视野。虽然人们生活在到处都有复杂混乱现象的现实世界里,科学家们看到的却只是钟表式的机械世界,科学的任务只是阐明这架钟表的结构和运行规律,而将不确定看作是无足轻重的,并将其排除在近代科学的研究对象范围之外。

随着统计力学、量子力学和非线性科学的深入发展,确定论思想在越来越多的研究领域受到了巨大的冲击。人们逐渐认识到不确定性有着普遍的来源,是客观世界固有的普遍特征。在客观世界中,不确定性与确定性是相互对立而又相互统一的^[5]。

5.2.2 不确定性知识

在日常生活中,我们所接触到的知识和信息,常常具备一定程度的不确定性,主要表现在信息与知识的不完全、不一致、不可靠和不精确方面。造成不确定性的根本原因是客观事物的复杂多变,而人的认识能力有限。在目前发展阶段,人们只能一方面力求减少不确定性,另一方面还得依靠带有不确定性的信息和知识来指导自己的行动^[1,2]。

随机性和模糊性是不确定性最基本的两个方面,这两个方面的不确定性与人们的认识和实践关系十分密切。随机性是在事件是否发生的不确定性中表现出来的条件的不确定性,事物本身的性态和类属是确定的;而模糊性却是事物本身性态与类属的不确定性。因此可以认为,随机性是一种外在的不确定性,而模糊性则是一种内在的不确定性。

随机性是由于事物及其运动的偶然性引起的。例如,掷骰子试验中,在相同的条件下,掷出骰子的结果有6种可能,试验结果受许多偶然因素影响,这种现象称为随机现象。其特点是:可重复观察,在观察之前知道所有可能的结果,但不知道到底哪个结果会出现。在这种情况下,条件与结果之间没有必然的因果关系,因而在事件的出现与否上表现出不确定性。

一方面,随机性使世界和人们的生活充满了未知的魅力,是创造性不可缺少的因素。另一方面,随机性也使得人类对客观世界的探索更为艰巨,科学家们认识世界时需要更复杂的理论。由概率论、数理统计和随机过程构成的概率理论,为研究随机性奠定了数学基础,也为研究不确定性提供了工具^[4]。

不确定性的早期研究内容仅仅是针对随机性,随着研究的深入,人们发现有一类不确定现象无法用随机性来描述,这就是模糊性。产生模糊性的原因,是因为事物在性态和类属上两极对立中有着丰富的中间过渡。

两极对立的中间过渡,表现在从一极到另一极存在着许多连续或离散的中间状态。例如,从人的身高来说,高与矮可以认为是两极,但人的身高从零点几米到两米多,几乎每一个高度的人都存在,不高不矮的中间状态是很多的,而这中间状态恰好反映了亦此亦彼的特点。一个中等身材的人,在一定条件下,在某些人眼里可能被认为是高个子,而在另外的条件下,在另一些人眼里却被认为是矮个子。再比如,健康人与不健康的人之间没有明确的划分,当判断某人是否属于“健康人”的时候,便可能没有确定的答案,这就是模糊性的一种表现。当一个概念不能用一个界限分明的集合来表达其外延的时候,便有某些对象在概念的正反两面之间处于亦此亦彼的状态,它们的类属划分便不分明了,呈现出模糊性。所以模糊性也就是概念外延的不分明性,事物对概念归属的亦此亦彼性。

模糊性还根源于事物的复杂性与发展变化性。例如,“天气好”涉及许多因素的交错作用,如温度、日照、风力和湿度等,而且这些因素又是在随时变化的,我们无法细致地一一描述,就只能笼统地做出一个“天气好”的结论。

模糊性是精确性的对立面,但不能消极地理解模糊性代表的是落后的生产力,恰恰相反,我们在描述客观事物时,经常借助于模糊性。早在1923年,著名的哲学家和数学家Russell就在一篇名为“含糊性”的论文中指出:几乎所有的自然语言均是模糊的,比如“红的”和“老的”等概念没有明确的内涵和外延,因而不明确的和模糊的^[6]。可是,在特定的环境中,人们用这些概念来描述某个具体对象时却又能心领神会,很少引起误解和歧义。美国著名的系统科学家Zadeh教授在1965年发表了著名的论文“模糊集合”,文中首次提出表达事物模糊性的重要概念:隶属度(或隶属函数),从而突破了康托尔的经典集合理论,奠定了模糊集理论的基础^[7]。

模糊性和随机性都是不确定性,模糊集理论使得区别于随机性的模糊性得到了一种数学表达,从而不确定性的知识表示和处理有了一套新的理论和方法。模糊集理论需要数据集合之外的先验信息,如要预先确定隶属度或隶属函数。一旦离开了隶属度或隶属函数,几乎所有的模糊集合运算都将难以进行。那么能否用不确定性本身提供的信息来研究知识的不确定性呢?1982年,波兰科学家Pawlak基于边界区域的思想提出了粗糙集的概念,宣告了粗糙集理论的诞生^[8]。

粗糙集理论是建立在知识分类机制的基础上的,它将分类理解为对象在特定空间上的等价关系,而等价关系构成了对该空间的划分,每一被划分的集合称为概念。粗糙集理论的主要思想是利用已知的知识库,将不精确或不确定的知识用已知的知识库中的知识来近似刻画。该理论与其他处理不确定和不精确问题的理论最显著的区别是它无须提

供数据集合之外的任何先验信息，所以对问题的不确定性的描述或处理可以说是比较客观的。粗糙集理论与概率理论和模糊集合理论等其他处理不确定或不精确问题的理论有很强的互补性。

5.3 确定性知识表示方法

知识可以分为确定性和不确定性两类，相应的知识表示方法也可以分为确定性知识表示方法和不确定性知识表示方法两类。确定性知识表示方法主要有逻辑表示法、关系表示法、产生式规则表示法、框架表示法、语义网络表示法、面向对象表示法和本体表示法等。

5.3.1 逻辑表示法

逻辑表示法是在人工智能领域中最先使用的一种知识表示方法^[9,10]，它使用命题演算（或称命题逻辑）、谓词演算（或称谓词逻辑）等来描述一些事实，并根据已有事实推出新事实。

首先介绍一下命题演算。命题演算研究由命题为基本构成单位的前提与结论之间的可推导关系。所谓命题，是指具有真假意义的陈述句。一个命题的真或假称为命题的真值（或简称为值），“真”用 T 表示，“假”用 F 表示。由于命题只有真、假两个值，所以命题逻辑也称二值逻辑。不能分解为更简单的陈述句，称为原子命题。

在自然语言中，常用“或”、“与”、“否则”等连接词把单个语句组合成复杂语句，在命题演算中，使用逻辑连接词来把单个（原子）命题组合成复合命题。常用连接词有五种：① 否定连接词（ \neg ）；② 合取连接词（ \wedge ）；③ 析取连接词（ \vee ）；④ 条件连接词（ \Rightarrow ）；⑤ 双条件连接词（ \Leftrightarrow ）。5 个基本的命题连接词在自然语言中通常分别用词语“非（不，没有）”、“并且（既……又……）”、“或者”、“如果……则”以及“当且仅当”表达。连接词都有从已知命题得到新命题的作用，从这个意义上讲，它们具有操作或运算的意义。在五个连接词中，否定词属一元连接词，其余 4 个都是连接两个（原子）命题以构成复合命题，称为二元连接词。

利用原子命题和五个基本的命题连接词可以构成复合命题，其真值与原子命题 A 和 B 的真值关系如表 5.1 所示。

表 5.1 真值表

A	B	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$	$\neg A$
T	T	T	T	T	T	F
T	F	F	T	F	F	F
F	T	F	T	T	F	T
F	F	F	F	T	T	T

复合命题的真值只取决于构成它们的各原子命题的真值，而与它们的内容、含义无关，与连接词所连接的两原子命题之间是否有关系无关。

在命题演算中，把命题分解到原子命题为止，并认为原子命题是不能再分解的，而仅仅研究以原子命题为基本单位的复合命题之间的逻辑关系和推理。虽然命题演算能够把客观世界的各种事实表示为逻辑命题，但是具有较大的局限性，即不适合用于表示比较复杂的问题。所以，在研究某些复杂问题的知识表示和推理时，有必要对原子命题作进一步分析，分析出其中的实体、谓词和量词，并研究它们的形式结构和逻辑关系、正确的推理形式和规则，这些正是谓词演算的基本内容。

下面再简单介绍一下谓词演算。一个原子命题由主语和谓词两部分组成。主语一般是可以独立存在的具体或抽象的实体，谓词用于刻画实体的性质或关系。

一般用大写字母表示谓词，小写字母表示实体名称（常量符号和变量符号）。例如，设 S ：是位大学生， c ：张明，则“张明是位大学生”可表示为 $S(c)$ 。这是一元谓词，因为只有一个实体。如果有两个实体 a 与 b ，而谓词 Q 表示“相等”，则“ a 等于 b ”可表示为 $Q(a, b)$ 。因为有两个实体，所以这是二元谓词。一般地，由一个谓词 P 和 n 个描述实体的变量符号 x_1, x_2, \dots, x_n 组成的 $P(x_1, x_2, \dots, x_n)$ 称为 n 元谓词。描述实体的变量符号的论述范围，称为论域。

n 元谓词的作用是表示出 n 个实体之间的关系。为了更加准确地表达某些命题，还需要引入用以刻画“所有的”、“存在一些”等表示不同数量的词，即量词。符号“ \forall ”称为全称量词符，用来表达“所有的”和“一切的”等词语；符号“ \exists ”称为存在量词符，用来表达“存在一些”和“至少有一个”等词语。有了量词符号后，用逻辑符号表示命题的能力大大加强了。

为了方便处理数学和计算机科学的逻辑问题及谓词表示的直觉清晰性，还需要引入项的概念。项可递归定义如下：① 单独的一个实体是项（包括常量和变量）；② 若 f 是 n 元函数符号，而 t_1, t_2, \dots, t_n 是项，则 $f(t_1, t_2, \dots, t_n)$ 是项；③ 任何项都由上述规则生成。下面给出原子公式的定义。若 $P(x_1, x_2, \dots, x_n)$ 是 n 元谓词， t_1, t_2, \dots, t_n 是项，则称 $P(t_1, t_2, \dots, t_n)$ 是原子公式。

原子公式是谓词演算的基本积木块，应用否定（ \neg ）、合取（ \wedge ）、析取（ \vee ）、条件（ \Rightarrow ）、双条件（ \Leftrightarrow ）等连接词，能够组合多个原子公式以构成比较复杂的合式谓词公式。

在使用谓词逻辑表示知识的时候，一般可以基于下面几步来进行：① 定义谓词及实体，确定每个谓词及实体的确切含义；② 根据所要表达的事物或概念，为每个谓词中的变量赋予特定的值；③ 根据所要表达的知识的语义，用适当的连接词将各个谓词连接起来，形成谓词公式。

5.3.2 关系表示法

一个关系可以认为是一个谓词外延的表示，而一个谓词反过来又可以是一个关系的内涵表示。根据这种对应关系，可以看出关系也是一种可以用来表示知识的方法。一个关系表面上看来就是一张二维表，表的每一行称为一个元组，每一列称为一个属性，一个元组就是该关系所涉及的属性集的笛卡儿积的一个元素。用表格形式来表示知识的例子很多，如航空公司的客运、货运价格表、国际和国内邮费表等。

用关系表示知识，就使人们有可能采用关系型数据库管理系统的功能来处理知识，用查询来代替推理，这样可以大大提高效率，而且有利于把数据库与知识库结合起来，把数据处理作为知识处理的一个有机组成部分。

用于数据库管理的关系模型是基于谓词逻辑和集合论的一种数据模型。关系数据模型是以集合论中的关系概念为基础发展起来的。关系模型中无论是实体还是实体间的联系均由单一的结构类型——关系来表示。在实际的关系数据库中的关系也称为表。一个关系数据库就是由若干个表组成的。

由于关系是元组的集合，因此关系模式要指出元组集合的结构。对关系的结构描述称为关系模式，它可以形式化地表示为 $R(U, D, \text{DOM}, F)$ ，其中 R 为关系名， U 为属性集， D 为域的集合， DOM 为属性向域映射的集合， F 为函数依赖的集合。关系模式相当于记录类型，是属性的有序集合。

关系实际上就是关系模式在某一时刻的状态或内容。关系模式是静态的、稳定的，而关系是动态的、随时间不断变化的，因为关系操作在不断地更新着数据库中的数据。但在实际当中，常常把关系模式和关系系统称为关系。常用的关系操作包括查询操作和插入、删除、修改操作两大部分。其中查询操作的表达能力最重要，它包括选择、投影、连接、除、并、交、差等。

关系模型中的关系操作能力早期通常是用代数方法或逻辑方法来表示的，分别称为关系代数和关系演算。关系代数是使用对关系的代数运算来表达查询要求的方式；关系演算是使用谓词来表达查询要求的方式。关系演算使用起来比关系代数容易得多，但是它仍然是基于逻辑学原理的，不容易被大多数人所掌握。因此，需要一个易于使用的关系演算工具，结构化查询语言（SQL, Structured Query Language）就是其中的一个，它巨大的成功使其成为了关系数据库模型领域影响力极大的语言。SQL 能够表达由关系演算支持的任何查询，因此在这个意义上，它被认为是一种“关系完整”的语言。

5.3.3 产生式规则表示法

产生式规则表示法是专家系统中应用最广泛的一种知识表示方法，主要原因在于产

生式规则最适合表示各种启发式的经验性规则，领域专家可以无须知识工程工具就能够把自己的知识转换成规则的形式。每一条产生式对应一条规则，最基本的形式为：IF a THEN b，其中 a 称为前件，b 称为后件。前件表示前提条件，各个条件由逻辑连接词（合取、析取等）组成各种不同的组合。后件表示当前提条件为真时，应采取的行动或所得的结论。大量产生式规则可以构成一个知识库，求解问题时将输入事实与各规则的前提对比，逐步进行推论。

例 5.1 动物识别系统 IDENTIFIER 的规则库。IDENTIFIER 是一个用于识别动物的分析系统，该系统利用产生式规则可以根据输入的动物特征来识别动物。现有猎豹、老虎、长颈鹿、斑马、鸵鸟、企鹅和海燕 7 种动物，以及如下 15 条判断规则：

- R1: IF 有毛发 THEN 是哺乳动物
- R2: IF 能产乳 THEN 是哺乳动物
- R3: IF 有羽毛 THEN 是鸟类动物
- R4: IF 能飞行 AND 能生蛋 THEN 是鸟类动物
- R5: IF 是哺乳动物 AND 吃肉 THEN 是食肉动物
- R6: IF 是哺乳动物 AND 有爪子 AND 有利齿 AND 眼睛前视 THEN 是食肉动物
- R7: IF 是哺乳动物 AND 有蹄 THEN 是有蹄动物
- R8: IF 是哺乳动物 AND 能反刍 THEN 是有蹄动物
- R9: IF 是食肉动物 AND 是黄褐色 AND 有深色斑点 THEN 是猎豹
- R10: IF 是食肉动物 AND 是黄褐色 AND 有黑色条纹 THEN 是老虎
- R11: IF 是有蹄动物 AND 有长腿 AND 有长颈 AND 是黄褐色 AND 有深色斑点 THEN 是长颈鹿
- R12: IF 是有蹄动物 AND 是白色 AND 有黑色条纹 THEN 是斑马
- R13: IF 是鸟类动物 AND 不会飞 AND 有长腿 AND 有长颈 AND 是黑色和白色相杂 THEN 是鸵鸟
- R14: IF 是鸟类动物 AND 不会飞 AND 能游水 AND 是黑色和白色相杂 THEN 是企鹅
- R15: IF 是鸟类动物 AND 善于飞行 THEN 是海燕

例如，输入的事实为“能飞行，能生蛋，善于飞行”，则可以判断该动物为海燕。以上 15 条规则中的条件和结论都可以分解为多个事实，如有毛发、是哺乳动物、是猎豹，但它们具备不同的属性，“有毛发”属于原始的事实，“是哺乳动物”是中间事实，“是猎豹”是结论性质的事实。故中间事实有：是哺乳动物，是鸟类动物，是食肉动物，是有蹄动物，结论性质的事实有：是猎豹，是老虎，是长颈鹿，是斑马，是鸵鸟，是企鹅，是海燕，其余均为原始事实。而每一个规则由两部分组成，即前提条件和结论，其中前提条件由若干条非结论性质的事实组成，结论由一条非原始性质的事实组成。

系统首先根据这 15 条规则和上述分析的事实与规则建立知识库，而后每一次输入事实集合的识别过程如下：① 采用正向推理，遍历知识库中所有的规则。对于某一条规则，判断输入事实集合中的事实是否能满足该规则的前提条件，若不满足则继续验证下一条

规则；若满足，则标记输入事实集中的满足前提条件的事实为已使用，再判断该规则的结论是否已经存在于输入事实集中，若不存在，则将规则的结论加入输入事实集合，并从第一条规则开始重新验证(只要输入事实集合在判断验证的过程中加入了新的事实，则需从第一条规则开始重新验证)；若存在，则继续验证下一条规则。② 遍历结束后，判断输入事实集合，若存在零个或一个以上的结论性质的事实，则该动物无法识别；若存在未被标记为已使用的事实，则该动物无法识别；若存在且仅存在一个结论性质的事实，则该动物被识别。

这种方法的优点是与人的思维方式接近，人们易于理解其内容，便于知识获取，规则具有独立性，易于修改和扩充；缺点是求解复杂问题时效率较低。

5.3.4 框架表示法

框架理论的基本观点是人脑已存储有大量的典型情景，当人们面临新的情景时，就从记忆中选择一个称作框架的基本知识结构，这个框架是以前记忆的一个知识空框，而其具体内容依照新的情景而改变，通过对这空框的细节加工修改和补充，形成对新情景的认识又记忆于人脑中。框架理论将框架视作知识的单位，将一组有关的框架连接起来形成框架系统。系统中不同的框架可以有共同的节点，系统的行为由系统内框架的变化来表现。推理过程是由框架间的协调来完成的。

框架通常由描述事物的各个方面的槽组成，每个槽可以拥有若干个侧面，而每个侧面又可以拥有若干个值^[10]。这些内容可以根据具体问题的具体需要来取舍，一个框架的一般结构如表 5.2 所示。

表 5.2 框架的一般结构

<框架名>			
<槽 1>	<侧面 11>	<值 111>
	⋮		
<槽 2>	<侧面 21>	<值 211>
⋮	⋮	⋮	
<槽 n>	<侧面 n1>	<值 n11>
	⋮		
	<侧面 nm>	<值 nm1>

较简单的情景是用框架来表示诸如人和房子等事物。例如，一个人可以用其职业、身高和体重等属性描述，因而可以用这些属性组成框架的槽。当描述一个具体的人时，再用这些属性的具体值填入到相应的槽中。下面给出的是一个简单的框架例子，如表 5.3 所示。

表 5.3 简单框架示例

John	
IS_A	Person
Profession	Programmer
Height	1.8m
Weight	79kg

对于大多数问题，不能这样简单地用一个框架表示出来，必须同时使用许多框架组成一个框架系统。图 5.1 给出了宾馆房间的框架描述。在框架系统中，各个框架之间通过 ISA 链表现了框架之间特殊与一般的继承关系。框架系统的层次结构不仅有利于查询和检索，而且由于可以避免存储一些重复的内容从而节省大量的存储空间。

框架的优点是对知识可以进行深化表达，因为它可用嵌套结构通过多方面、多重属性对事物进行描述；缺点是框架内部不具备推理所需的一些规则。

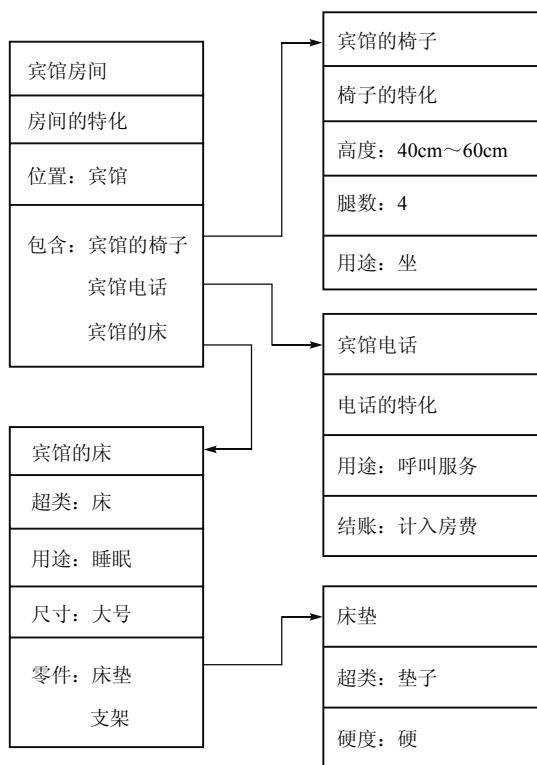


图 5.1 宾馆房间的框架描述

5.3.5 语义网络表示法

语义网络是知识表示中最重要方法之一，是一种表达能力强而且灵活的知识表示

方法。语义网络是由节点和连接节点的边所组成的有向图。其中的节点表示对象、概念、事件和行为等，边表示节点间的关系。由于关系可以是各式各样的，这就使语义网络的表示能力大大增强。图 5.2 是一个简单的语义网络的例子。

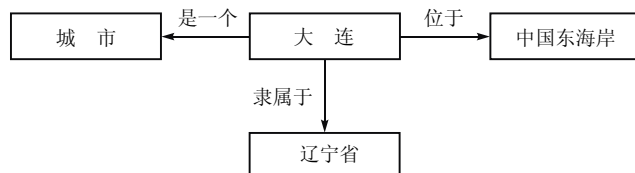


图 5.2 一个简单的语义网络的例子

采用语义网络表示法比较适合领域大多数是根据非常复杂的分类进行推理的领域以及需要表示事件状况、性质以及动作之间的关系的领域。语义网络表示法的优点是表示方式直观，易于理解，可把实体间的结构、属性、因果关系简单明了地表达出来，有利于认识的深化。它的缺点是节点所表示的概念没有意义上的约定，一些非线性，网络状、递归的联系会使存储与检索效率降低。

5.3.6 面向对象表示法

面向对象方法是以对象作为基本的逻辑构造；用类描述具有相同特征的一组对象，整体地代表这类对象；以继承作为共享机制，共享类中的方法和数据；对象之间则以消息传递的方式进行通信^[12]。因此人们经常将面向对象方法描述为如下的一个公式

$$\text{面向对象} = \text{对象} + \text{类} + \text{集成} + \text{消息通信}$$

用面向对象方法表示知识时，对象可以是某种知识，它包括与知识有关的属性及操作知识的方法；类是对多种知识的抽象组织，是一类知识共性的集合，每一个知识对象都是对应类的实例；继承是类的组织方法；类与类之间可以通过继承来共享知识，表示领域之间的从属关系；消息通信是各知识对象间相互联系的方式，通过消息通信可利用对象的属性和方法来对知识进行操作。

根据知识的波粒二象性，可将知识表示方法分为说明性表示和过程性表示两类。说明性表示只给出事物本身的属性及事物之间的相互关系，对问题的解答就隐含在这些知识之中；而过程性表示则给出解决一个问题的具体过程。两者相比，说明性表示比较简要、清晰、可靠和便于修改，但往往效率低；过程性表示则与此相反，它比较直截了当，效率高，但由于详细地给出了解决问题的过程，使得这种知识表示显得复杂、不直观、易出错和不便修改。在知识表示中引入面向对象技术，恰好能够很好地将知识的说明性与过程性有机地结合起来，是复杂知识表示的一种较好的方法。

在基于面向对象的知识表示中，知识可看作是对象，对象的属性是对知识对象的说

明性表示,对象的方法是对知识对象的过程性表示。不同特性的知识可抽象成类,类有子类,子类还有子类,如此形成一个层次体系。同一子类中所有对象实例具有共同属性,均使用同样的方法。这些属性和方法顺着子类的层次体系实行继承。知识对象的方法通过消息的传递而被调用。

知识以面向对象技术表示之后,可得到知识类的继承关系层次体系结构,根据这种结构,便可构造知识库结构。

5.3.7 本体表示法

本体是对领域实体存在本质的抽象,它强调实体间的关联,并通过多种知识表示元素将这些关联表达和反映出来,这些知识表示元素也被称为元本体,主要包括:

(1) 概念——表示领域知识元,包括一般意义上的概念以及任务、功能、策略、行为和过程等,在本体的实现中,概念通常用类来定义,而且通常具有一定的分类层次关系。

(2) 属性——描述概念的性质,是一个概念区别于其他概念的特征,通常用槽或者类的属性来定义。

(3) 关系——表示概念之间的关联,如一些常用的关联有父关系、子关系、相等关系等。

(4) 函数——表示一类特殊的关系,如长方形的长和宽唯一决定其面积。

(5) 公理——表示永真式,在本体中,对于属性、关系和函数都具有一定的关联和约束,这些约束就是公理,公理一般用槽的侧面来定义。

(6) 实例——表示属于某个概念类的具体实体。

构造本体的目的是为了实现某种程度的知识共享和重用,本体的作用主要有以下两方面:① 澄清了领域知识的结构,可以重用,从而避免了重复的领域知识分析;② 统一的术语和概念使知识共享成为可能。

本体作为一种知识表示方法,与谓词逻辑、框架等其他方法的区别在于它们属于不同层次的知识表示方法,本体表达了概念的结构、概念之间的关系等领域中实体的固有特征,即共享概念化,而其他的知识表示方法如语义网络等,可以表达某个体对领域中实体的认识,但不一定是实体的固有特征。这正是本体层与其他层次的知识表示方法的本质区别。

知识工程师将本体概念引入知识工程,详细说明模型中涵盖的概念、实例、关系和公理等实体,并以此建立本体。通过使用元属性对属性进行分析,并对属性提出了一种针对本体建模概念化分析的形式化方法,解决了知识共享中的一些问题,有效地促进了来自不同领域的研究人员与组织间的交流。基于本体的知识表示法在知识表示方面有很大的发展潜力。

5.4 不确定性知识表示方法

不确定性知识表示方法主要有基于概率、基于模糊集和基于粗糙集等不确定数学理论的知识表示方法。

5.4.1 基于概率论的知识表示法

随机性是一种最常见的不确定性，处理这种不确定性的数学工具是由概率论、数理统计和随机过程构成的概率理论。概率论研究和处理随机现象，事件本身有明确的含义，只是由于条件不充分，使得在条件与事件之间不能出现决定性的因果关系。

确定某一事件 A 的概率 $P(A)$ 通常有三种计算方法，即古典概率法、频率法和主观概率法。在古典概率法中， $P(A) = k/n$ （其中 k 为 A 中所包含的基本事件数， n 为基本事件的总数）；在频率法中， $P(A) = m/n$ （其中 n 为重复试验次数， m 为事件 A 出现的次数）；在主观概率法中， $P(A)$ = 专家主观赋值（通常用于不宜大量重复的随机现象）。

贝叶斯定理是概率论中的一个著名定理，这一定理可用一个数学公式来表达，即贝叶斯公式。假定 B_1, B_2, \dots, B_n 是某个过程的若干可能的前提，则 $P(B_i)$ 是人们事先对各前提条件出现可能性大小的估计，称为先验概率。如果这个过程得到了一个结果 A ，那么贝叶斯公式

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{i=1}^n P(B_i)P(A|B_i)} \quad (5.1)$$

提供了根据 A 的出现而对前提条件做出新评价的方法。 $P(B_i / A)$ 就是对以 A 为前提下 B_i 的出现概率的重新认识，称 $P(B_i / A)$ 为后验概率。贝叶斯公式给出了先验概率与后验概率之间的关系，它能确定在产生结果 A 的各种原因中，哪一个在起着更重要的作用。例如，在医疗诊断中，设症状 A 可能由疾病 B_i 引起，为了诊断有症状 A 的人到底患了哪种疾病，可用贝叶斯公式计算在 A 出现的情况下，各疾病 B_i 发生的后验概率，然后按照后验概率的大小判断患者患何种疾病的可能性最大。

经过多年的发展与完善，贝叶斯公式及由此发展起来的一整套理论与方法，被统称为“贝叶斯方法”。贝叶斯方法是一种非常有代表性的不确定性知识表示和推理方法。

在贝叶斯方法中，由于全联合概率公式假设所有变量之间都具有条件依赖性，其计算复杂，在设计模式识别系统时常采用朴素贝叶斯分类器的简化形式。但是朴素贝叶斯分类器假设所有变量之间都是条件独立的，与实际不是很相符。而贝叶斯网络充分利用了变量之间的独立性和条件独立性关系，大大减小了为定义全联合概率分布所需要指定

的概率数目,同时也避免了朴素贝叶斯分类器要求所有变量都是独立的不足,是一个很好的折中办法。作为一种特殊的建模方式,贝叶斯网络在各领域的应用也越来越广泛。

贝叶斯网络是一种概率网络,用于表示变量之间的依赖关系,是带有概率分布标注的有向无环图,能够图形化地表示一组变量间的联合概率分布函数。

贝叶斯网络模型结构由随机变量(可以是离散或连续)集组成的网络节点、具有因果关系的网络节点对的有向边集合和用条件概率分布表示节点之间的影响(即权重集合)等组成。其中节点表示了随机变量,是对过程、事件、状态等实体的某些特征的描述;边则表示变量间的概率依赖关系。起因的假设和结果的数据均用节点表示,各变量之间的因果关系由节点之间的有向边表示,一个变量影响到另一个变量的程度用数值权重形式描述。

作为一种图形化的建模工具,贝叶斯网络具有以下五个特性。

(1) 贝叶斯网络将有向无环图与概率理论有机结合,不但具有正式的概率理论基础,同时也更具有直观的知识表示形式。一方面,它可以将人类所拥有的因果知识直接用有向图自然直观地表示出来;另一方面,也可以将统计数据以条件概率的形式融入模型。这样贝叶斯网络就能将人类的先验知识和后验的数据无缝地结合起来,克服框架、语义网络等模型仅能表达处理定量信息的弱点和神经网络等方法不够直观的缺点。

(2) 贝叶斯网络与一般知识表示方法不同的是对问题域的建模。因此当条件或行为等发生变化时,不用对模型进行修正。

(3) 贝叶斯网络可以图形化表示随机变量间的联合概率,因此能够处理各种不确定性信息。

(4) 贝叶斯网络中没有确定的输入/输出节点,节点之间是相互影响的,任何节点观测值的获得或者对于任何节点的干涉,都会对其他节点造成影响,并可以利用贝叶斯网络推理来进行估计预测。

(5) 贝叶斯网络的推理是以贝叶斯概率理论为基础的,不需要外界任何推理机制,不但具有理论依据,而且将知识表示与知识推理结合起来,形成统一的整体。

5.4.2 基于模糊集理论的知识表示法

知识的不确定性,还反映在语言的不确切性上,因为语言是知识的重要载体^[4]。自然语言中的基本单元是语言值,对应一个个概念。概念的不确切性有多个方面,最主要的不确切性就是模糊性。

人的自然语言,无论在语音、语义和语法诸方面,都存在模糊性,在人类文化发展过程中,科学技术的发展要求语言向精确化方向发展,以便精确地描述科学现象和规律;另一方面,模糊化使语言更丰富多采,更富有表现力,在文学艺术上有它特有的功能^[1]。

以数学语言和计算机程序设计语言为代表的形式化人工语言则是严格、精确的。这样一来,在自然语言与人工语言之间就存在一个鸿沟。利用模糊集理论,是否能将两者加以沟通,是人们关注的问题。这在知识应用和创新的活动中尤其重要。因为在知识创新活动中的许多事件,目前只能用自然语言描述,怎样把它与计算机系统上的可以形式化的数值、符号描述联系起来,是开拓计算机应用的重要环节,也是使计算机的工作更向人的思维靠近的重要步骤。因为人胜过机器的地方,如控制论创始人维纳(Wiener)所指出的,正在于“人有运用模糊概念的能力”。

概念通常用内涵和外延来表述,概念的模糊性主要指外延的不确定。我们把具有模糊概念的语言叫做模糊语言。关于语词处理,Zadeh提出了语言变量这一概念,它揭示了语词之间的系统联系,有利于分析语词的结构。

以人的年龄描述为例,如果甲今年20岁,乙今年70岁,丙今年55岁,这时可用数值描述,叫做数值变量。但在日常生活中,我们不说具体岁数,而说甲年轻,乙很老,丙有点老了,这时年龄仍是一个变量,但用语词“年轻”、“很老”为值,所以叫做语言变量,其值称为语言变量的语言值。语言变量是由人类的语言实践中提炼出来的概念,除了语言变量和语言值外,还需要语法规则与语义规则。使用语言变量去描述事物实际上相当于通过“粗粒化”操作进行了数据压缩。

在人们的工作经验中,往往也有许多模糊的东西。例如,要确定一炉钢水是否已经炼好,除了要知道钢水的温度、成分比例和冶炼时间等精确信息外,还需要参考钢水颜色、沸腾情况等模糊信息。人与计算机相比,一般来说,人脑具有处理模糊信息的能力,善于判断和处理模糊现象。但计算机对模糊现象识别能力较差,为了提高计算机识别模糊现象的能力,就需要把人们常用的模糊语言设计成机器能接受的指令和程序,以便机器能像人脑那样简洁灵活地做出相应的判断,从而提高自动识别和控制模糊现象的效率。这样,就需要寻找一种描述和加工模糊信息的数学工具。模糊化的基本数学工具还是模糊集理论。

集合是描述人脑思维对整体性客观事物的识别和分类的数学方法。康托尔集合论要求其分类必须遵从形式逻辑的排中律,论域(即所考虑的对象的全集)中的任一元素要么属于集合 A ,要么不属于集合 A ,两者必居其一,且仅居其一。这样,康托尔集合就只能描述外延分明的“明确概念”,只能表现“非此即彼”,而对于外延不分明的“模糊概念”则不能反映。在模糊集合中,给定范围内元素对它的隶属关系不一定只有“是”或“否”两种情况,而是还存在许多中间过渡状态。

设 U 是论域,称映射 $A(x): U \rightarrow [0,1]$ 确定了一个 U 上的模糊集合 A ,映射 $A(x)$ 称为 A 的隶属函数,它表示 x 对 A 的隶属程度(简称隶属度)。使 $A(x) = 0.5$ 的点 x 称为 A 的过渡点,此点最具模糊性。当映射 $A(x)$ 只取0或1时,模糊集合 A 就是经典集合,而 $A(x)$

就是它的特征函数。可见经典集合是模糊集合的一种特殊情形。

模糊集的常用表示法有下述几种：① 解析法，即给出隶属函数的具体表达式。② Zadeh 记法，如 $A=1/x_1+0.5/x_2+0.7/x_3+0/x_4$ 。分母是论域中的元素，分子是该元素对应的隶属度。③ 序偶法，如 $A=\{(x_1,1), (x_2,0.5), (x_3,0.7), (x_4,0)\}$ ，序偶对的前者是论域中的元素，后者是该元素对应的隶属度。④ 向量法，在有限论域的场合，若给论域中的元素规定一个表达的顺序，那么可以将上述序偶法简写为隶属度的向量式，如 $A=(1, 0.5, 0.7, 0)$ 。

例 5.2 设论域 $U=\{x_1(140), x_2(150), x_3(160), x_4(170), x_5(180), x_6(190)\}$ （单位：cm）表示人的身高，那么 U 上的一个模糊集“高个子” A 的隶属函数 $A(x)$ 用解析法可定义为

$$A(x)=\frac{x-140}{190-140}$$

也可用 Zadeh 记法表示为

$$A=\frac{0}{x_1}+\frac{0.2}{x_2}+\frac{0.4}{x_3}+\frac{0.6}{x_4}+\frac{0.8}{x_5}+\frac{1}{x_6}$$

用序偶法可表示为 $A=((x_1,0), (x_2,0.2), (x_3,0.4), (x_4,0.6), (x_5,0.8), (x_6,1))$ ，还可向量法表示为 $A=(0, 0.2, 0.4, 0.6, 0.8, 1)$ 。另外，还可以在 U 上建立“矮个子”、“中等个子”等模糊子集。

隶属函数是模糊数学中最重要的概念之一，模糊集合的一个基本问题就是如何确定一个明晰的隶属函数，但至今尚没有严格的确定方法，通常靠直觉、经验、统计、排序和推理等方式确定。

由于模糊性概念可以用模糊集合来描述，人们运用模糊概念进行判断、评价、推理、决策和控制的过程也可以用模糊数学的方法来描述。例如，模糊聚类分析、模糊模式识别、模糊综合评判、模糊决策与模糊预测、模糊控制和模糊信息处理等。计算机智能信息处理是模糊数学最重要的应用领域之一，模糊数学已经被用于专家系统和知识工程等方面，在很多领域中发挥着非常重要的作用，并已获得巨大的经济效益。

5.4.3 基于粗糙集理论的知识表示法

粗糙集理论是一种描述不精确、不确定性知识的数学工具。粗糙集理论从知识分类入手研究不确定性。“知识”这个概念有多种不同的含义。在粗糙集理论中，“知识”被认为是一种分类能力。而人们的行为是基于分辨现实的或抽象的对象的能力。例如，在远古时代，人们为了生存必须能分辨出什么可以食用，什么不可以食用；医生给病人诊断，必须辨别出患者得的是哪一种病。这些根据事物的特征差别将其分门别类的能力均

可以看作是某种“知识”。基于粗糙集的知识表示思想可以概括为：一种类别对应于一个概念，知识由概念组成；如果某知识中含有不精确概念，则该知识不精确；对不精确概念的描述通过下近似（Lower Approximation）概念和上近似（Upper Approximation）概念来表示。

与模糊集理论类似，粗糙集理论也拓展了经典的集合论，把用于分类的知识嵌入集合内，作为集合组成的一部分。一个对象 a 是否属于集合 X 需根据现有的知识来判断，可分为三种情况：① 对象 a 肯定属于集合 X ；② 对象 a 肯定不属于集 X ；③ 对象 a 可能属于也可能不属于集合 X 。集合的划分是相对的而不是绝对的，密切依赖于我们所掌握的关于论域的知识。

给定一个有限的非空集合 U 称为论域， R 为 U 中的一族等价关系，即关于 U 的知识，则二元组 $K=(U, R)$ 称为一个近似空间（Approximation Space）。设 x 为 U 中的一个对象， X 为 U 的一个子集， $R(x)$ 表示所有与 x 不可分辨的对象所组成的集合，换句话说，是由 x 决定的等价类，即 $R(x)$ 中的每个对象都与 x 有相同的属性。集合 X 关于 R 的下近似定义为

$$\underline{R}(X) = \{x \in U \mid R(x) \subseteq X\} \quad (5.2)$$

式中， $\underline{R}(X)$ 实际上由那些根据现有知识判断肯定属于 X 的对象所组成的最大的集合，有时也称为 X 的正区（Positive Region），记作 $\text{POS}(X)$ 。类似地，由根据现有知识判断肯定不属于 X 的对象组成的集合称为 X 的负区（Negative Region），记作 $\text{NEG}(X)$ 。集合 X 关于 R 的上近似定义为

$$\bar{R}(X) = \{x \in U \mid R(x) \cap X \neq \emptyset\} \quad (5.3)$$

式中， $\bar{R}(X)$ 是由所有与 X 相交非空的等价类 $R(x)$ 的并集，是那些可能属于 X 的对象组成的最小集合。显然，有

$$\bar{R}(X) + \text{NEG}(X) = U \quad (5.4)$$

集合 X 的边界区域（Boundary Region）定义为

$$\text{BND}(X) = \bar{R}(X) - \underline{R}(X) \quad (5.5)$$

式中， $\text{BND}(X)$ 为集合 X 的上近似与下近似之差。如果 $\text{BND}(X)$ 是空集，则称 X 关于 R 是精确的；反之如果 $\text{BND}(X)$ 不是空集，则称集合 X 为关于 R 的粗糙集。

粗糙集中，集合（概念）的不精确性是由于边界区域的存在而引起的。集合的边界区域越大，精确性就越低。粗糙集用引入精度的概念来表达这一点，精度越大，精确程度越高。由等价关系 R 定义的集合 X 的精度定义为

$$d_R(X) = \frac{\#\underline{R}(X)}{\#\bar{R}(X)} \quad (5.6)$$

式中,“#”表示集合的基数或势,对有限集合则表示集合中所包含的元素个数。显然 $0 \leq d_R(X) \leq 1$, 如果 $d_R(X)=1$, 则称集合 X 相对于 R 是清晰的; 如果 $d_R(X)<1$, 则称集合 X 相对于 R 是粗糙的。

粗糙集和模糊集都能处理不完备数据,但方法不同。模糊集注重描述信息的含糊程度,粗糙集则强调数据的不可辨别、不精确和模棱两可。使用图像处理中的语言来作比喻,当论述图像的清晰程度时,粗糙集强调组成图像像素的大小,而模糊集则强调像素存在不同的灰度。粗糙集研究的是不同类中的对象组成的集合之间的关系,重在分类;模糊集研究的是属于同一类的不同对象的隶属关系,重在隶属的程度。因此粗糙集和模糊集是两种不同的理论,但又不是相互对立的,它们在处理不完备数据方面可以互为补充。

5.5 非结构化信息的知识表示方法

随着全球信息化进程的加快,非结构化信息的处理正逐渐成为当今信息处理的主题,而非结构化信息的知识表示方法是与之相关的核心科学问题之一,近年来也取得了重要的进展。

5.5.1 半结构化数据与非结构化数据

数据是信息和知识的重要载体,它通常是指人类通过不同的传感方式所获得的原始资料,如表格、曲线、图形、文字、网页、电子邮件、文本、图像和视频等。能够用有限规则完全表征与刻画、且在可接收时间内可以形式化处理的数据称为结构化数据,反之则称为非结构化数据^[13]。表格、曲线、几何图形、有限规则集和程序等是结构化数据;而由于语义的不确定性,文本是非结构化数据;由于视觉感知的主观依赖性和歧义性,图像和视频等是非结构化数据。网页、电子邮件是半结构化数据,半结构化数据也是一种非结构化数据,一般指介于严格结构化的数据(如关系数据库中的数据)和完全无结构的数据(如图像和视频文件)之间的数据形式。半结构化数据具有一定的结构,但其结构与数据混在一起,没有明显的结构模式定义。例如,HTML(Hyper Text Markup Language)文件,它的标签使得文档具有一定的结构,但标签与数据混在一起,是一种隐性结构,即结构模式与数据间的界限混淆。

非结构化数据是远比结构化数据广泛存在的数据形式。随着传感技术、网络技术和计算机技术的迅速发展及普及,异构分布式传感网络与互联网已经广泛应用,从外部世界获取非结构化数据(信息)已变得极为便利与快捷。无论从信息的获取手段,还是信息存储方式,都使得非结构化数据已无处不在。从可见光视频摄像机到多光谱成像设备,从脑电阵列设备到相控阵雷达,从单传感器到分布式传感网络,从个人计算机到互联网及万维网,人类所面对的已经是一个五彩绚丽、复杂纷呈的非结构化数据海洋。这些非

结构化数据常表现出以下共有特征：形式上的高维、海量、异构与动态；内容上的不完整、不确定、无序与歧义；表达上难以用有限规则刻画；解译与应用上依赖信息利用主体的感知和理解等。

由于感知的主观依赖性和歧义性，文本、图像和视频是典型的非结构化信息，而以非结构化信息为基本对象，研究非结构化信息的编码、分析、解译和利用等是当前智能信息处理的前沿科学问题^[13]。

非结构化信息处理已逐渐成为当今信息处理的主题。非结构化信息所具有的动态、异构、混杂和高维等特征显著增加了知识表示的困难性和推理的复杂性，如何形式化这类信息并进行有效推理已成为当前最具挑战性的研究课题和信息技术发展的新瓶颈。与之相关的核心科学问题包括：① 非结构化信息的知识表示与建模，即如何形式化表示或编码具有多层次的混合数据和信息，以及建立可物理实现或可编程的计算模型，这是解决非结构化数据和信息有效利用的前提；② 非结构化信息的数学建模与知识推理，特别是如何从非结构化数据中提取具有物理意义的信息原子，将非结构化信息转化为结构化信息，进而建立有序的信息链，这是解决信息精化和知识提取的核心；③ 非结构化信息的综合利用与行为决策，特别是如何对异构、跨媒体、不同粒度和层次上的信息（如传感数据）有效综合，以建立综合集成、优势互补的行为决策机制，这是提升非结构化信息应用水平与决策的关键。

从非结构化数据中抽取有用的信息和知识，建立简洁有效的信息链，传统的信息处理和分析技术面临巨大挑战。

5.5.2 半结构化数据的知识表示

Web 上的大量数据是半结构化的。Web 半结构化数据的分布有三种形式：① Web 网页，即以 HTML 描述格式和组织 Web 网页，以及通过 XML (eXtensible Markup Language) 进行数据交换和组织。② Web 网站，包括网页内容、锚文本和链接关系等导航信息。③ Web 网络。搜索引擎是 Web 网络上最常用的工具之一，用以提供对某一主题的高效检索。

Web 半结构化数据分布于网页、网站和网络中，且在 Web 服务器上，包含于 HTML 格式等文件中，要从这样的文件中抽取和描述半结构化数据，必须提供一种说明文件，建立一种知识表示与抽取模型，用于指导程序自动识别所需抽取的值，并加上标记信息。半结构化数据语义结构信息不够完整，不利于数据的深层次处理与分析。因此，Web 挖掘首先就需要解决异构数据的集成和查询问题，这就需要构造半结构化数据知识表示和抽取模型。

对象交换模型 (OEM, Object Exchange Model) 是一种自描述对象模型，专为表示

半结构化数据而设计,在异构数据源的集成应用中非常有用。它可被看成是一个图,其中的每个顶点表示一个数据对象,而依附于两个顶点之间的边的标号表示两个顶点(对象和其子对象)之间的层次引用。OEM 图中的每个数据对象具有两个属性:标号(Identifier)和值(Value)。对象的标号唯一地标识了该对象,而对象的值或者是原子型(如整数或字符串),或者是对象引用集合(记作一组〈lable, id〉对组成的集合)。具有原子值的对象被称为原子对象,而具有集合值的对象被称为复合对象。可以用 Identifier(o)和 Value(o)表示对象 o 的标号和值,用 object(id)唯一表示具有标识 id 的对象。

由于对象交换模型的简洁性和灵活性,因而可表示许多类型数据,包括关系数据及 HTML 和 XML 格式的电子文档等。OEM 模型可描述半结构化数据的整体结构,通过数据模型表示数据,以便从中找到公共结构,形成模式,使不同结构的数据源之间能够方便地进行数据交换或转换。

5.5.3 多媒体信息的知识表示

多媒体数据库是指存储和管理大量多媒体对象的数据库,如音频数据、视频数据、图像数据、文本数据和超文本数据。在网络时代,多媒体数据广泛存在于各个领域。

与关系数据库中的数据相比,多媒体数据具有如下四个特点。

(1) 多媒体数据复杂。从非结构化或半结构化的多媒体数据中抽取隐藏的知识或其他非显形存储的模式比较困难,且数据量大,更新变化快。

(2) 多媒体信息语义关联性强。这是因为各自独立的对象内蕴涵极为丰富的语义联系,所以进一步加大了多媒体信息处理的难度。

(3) 多媒体信息具有时空相关性。由于多媒体信息,尤其是声音、动画和视频等连续性媒体一般都有时间敏感性、空间相关性等特点,因此难以在多媒体信息和数据描述之间建立简单的对应关系,其信息特征的提取比较困难。

(4) 知识的表达和解释比较困难,这是因为多媒体信息处理所得出的模式往往比较隐晦。

多媒体数据包含了大量的语义和视听特征,这些特征都是非结构化的,提取这些特征并进行规范化描述和表示是多媒体信息处理的基础。一般而言,对于图像数据可以用颜色、纹理、形状和运动向量等来表示图像的基本特征;而视频由一系列图像帧组成,因此从帧的角度看,针对图像的内容分析和特征提取技术可以扩展地应用到视频分析中去。但是,仅仅使用这种扩展技术是不够的,因为视频的时空特性,使得视频包含比图像更丰富的内容,如故事情节、事件和行为等。因此需要在空间和时间两个方面对视频内容进行描述和分析。对于视频来说,还应该包括在视频结构特征和运动特征等方面的处理和特征提取。对于音频数据则主要提取声强、频率、谐波结构、音调、音色和音量

等特征。对于文本数据来说,主要提取描述特征(如文本的名称、日期、大小和类型等)和语义特征(如文本的作者、机构、标题和内容等)。这里要注意的是,描述特征易于获取,而语义特征则较难得到。近年来,万维网联盟(W3C, World Wide Web Consortium)指定的XML、RDF(Resource Description Framework)等规范提供了对Web文档资源进行描述的语言和框架。在此基础上,可以从半结构化的Web文档中提取作者和机构等特征。文本的内容特征提取的方法主要是基于词频和位置,即考虑到特征发生的频率和所处的位置。利用文本数据的特征表示,可以将其转化为结构化数据,向量空间模型是最常用的文本表示模型。

对于多媒体数据,利用内容处理技术提取多媒体对象和数据的时空、运动和视听特征以及对应多媒体结构元素的语义之后,还需要把这些特征有效地组织成为特征数据库,以支持后续的多媒体信息处理与挖掘工作。目前,多媒体信息处理仍是一个新的研究方向,一些概念和方法正在形成中,有很多问题有待进一步研究。



思考与讨论题

- (1) 确定性知识表示方法的要点是什么?各种方法之间有何关系?
- (2) 不确定性知识表示方法主要有哪几种?有何特征?
- (3) 与结构化数据相比,非结构化数据和信息有何特点?如何对非结构化数据和信息进行知识表示?



参考文献

- [1] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [2] 王众托. 知识管理[M]. 北京: 科学出版社, 2009.
- [3] 马鸣远. 人工智能与专家系统导论[M]. 北京: 清华大学出版社, 2006.
- [4] 李德毅, 杜鹑. 不确定性人工智能[M]. 北京: 国防工业出版社, 2005.
- [5] 普利高津. 确定性的终结[M]. 湛敏译. 上海: 上海科技教育出版社, 1998.
- [6] Russell B. Vagueness[J]. Australasian Journal of Philosophy and Psychology, 1923, 1: 84-92.
- [7] Zadeh L A. Fuzzy sets[J]. Information Control, 1965, 8: 338-353.
- [8] Pawlak Z. Rough set[J]. International Journal of Computer and Information Science, 1982, 11: 341-356.
- [9] 李盘林, 李丽双, 李洋, 王春立. 离散数学[M]. 北京: 高等教育出版社, 1999.

- [10] 蔡自兴, 徐光祐. 人工智能及其应用[M]. 北京: 清华大学出版社, 2004.
- [11] 史忠植, 王文杰. 人工智能[M]. 北京: 国防工业出版社, 2007.
- [12] 向阳, 王征, 于长锐. 复杂问题决策支持模式研究[M]. 北京: 科学出版社, 2008.
- [13] 徐宗本, 张讲社. 基于认知的非结构化信息处理: 现状与趋势[J]. 中国基础科学, 2007, 6: 4-8.

第 6 章

信息与知识的组织

内容提要

随着网络信息资源的爆炸式增长，传统的以图书情报资源为对象的信息组织方法已经不能满足人们对信息资源的有效获取和利用，因此，出现了知识组织方法，以此解决信息泛滥和信息垃圾问题。本章首先介绍信息组织过程中的几类方法，然后重点介绍知识组织的方法，包括语义词典、概念图、主题图、本体、语义网、特定领域分类法、网络分类法和分众分类法。

本章重点

- 信息组织的方法
- 知识组织的内涵
- 知识组织的方法



6.1 信息组织

为实现信息从无序到有序的转换，信息组织的方法和工具是必不可少的。下面就从信息组织的概念和类型出发，介绍不同活动中信息组织的方式和方法。

6.1.1 信息组织的概念

信息组织即信息序化或信息整序，序是事物的一种结构形式，是指事物或系统的各个结构要素之间的相互关系以及这种关系在时间和空间中的表现。当事物结构要素具有某种约束性且在时间序列和空间序列呈现某种规律性时，这一事物就处于有序状态；反之，则处于无序状态。

信息组织是利用一定的规则、方法和技术对信息的外部特征和内容特征进行揭示和描述，并按给定的参数和序列公式排列，使信息从无序集合转换为有序集合的过程。所谓信息的外部特征是指信息的物质载体所直接反映的特征，构成信息载体的外在的、形式的特征，如信息的物理形态、题名和责任者及信息的类型、信息生产和流通等方面的特征；信息的内容特征则是指信息所包含和承载的具体内容，即通过信息载体所传递和交流的具体内容。信息组织的基本对象和管理依据就是信息的外在特征和内容特征这两个方面^[1]。

6.1.2 信息组织的类型

信息组织的类型按信息表现形式，可划分为文字信息组织、图像信息组织、声音信息组织和视频信息组织。

信息组织的类型按信息的加工程度，可划分为一次信息组织、二次信息组织和三次信息组织。

信息组织的类型按信息的传播载体，可划分为文献信息源和非文献信息源。非文献信息源特指网络环境下没有以传统文献载体形式出现的信息源，如程序代码、网页和超文本等。

6.1.3 信息组织的原则

为避免信息工作的随意性、无计划性和盲目性等现象的出现，在对信息进行组织时，要遵循以下四个原则^[1]。

1. 客观性原则

在信息组织中用于描述和揭示的基本依据就是信息本身，因此，在描述和揭示信息外

在特征和内容特征时必须客观而准确,要根据信息本身所反映的各种特征加以科学的反映和序化,形成相应的信息组织的成果。信息组织的客观性原则还要求我们不断跟踪信息源的发展变化和信
息组织技术的发展变化,使信息组织与条件变化和环境变化保持客观一致性。

2. 系统性原则

为实现信息组织的系统性,必须把握四个关系,即① 宏观信息组织和微观信息组织的关系;② 信息组织部门与其他部门的关系;③ 信息组织工作的各个环节之间的关系;④ 不同信息处理方法之间的关系。

3. 目的性原则

信息组织具有鲜明的目的性,必须充分围绕用户的信息需求开展工作,必须充分注意信息机构目标市场的需求状态及其变化特征。此外,为了实现信息组织的目标还必须注意信息工作的计划性和长期性,以及其与信息机构本身性质、特点和能力的适应性。

4. 现代化原则

信息组织的现代化原则主要表现在两个方面,一是信息组织的思想观念现代化,主要体现在标准化方面,包括信息组织工作的统一性,信息组织方法的规范性,信息组织系统的兼容性和信息组织成果的通用性;二是信息组织的技术手段现代化,主要体现在自动化方面,包括自动标引、二次信息自动生成和数据库建设。

6.1.4 信息组织的方法

信息组织通常包括优化选择、确定标识、组织排序、改编重组四种活动,每一活动中,信息组织的方式方法各有不同^[1]。

1. 优化选择

优化选择是对初选信息的鉴别、筛选和剔除,是对社会信息流的进一步过滤和深层次的控制,其主要任务是去粗取精、去伪存真,使信息流具有更强的针对性和实效性。

1) 比较法

比较就是对照事物,以提示它们的共同点和差异点。通过比较,判定信息的真伪,鉴别信息的优劣,从而排除虚假信息,去掉无用信息,这是择优的基本方法。运用比较法,首先应找出事物可比的基础,即比较对象的可比事项。信息的可比事项包括时间、空间、来源和形式等。

2) 分析法

通过对信息内容的分析而判断其正确与否、质量高低和价值大小等。例如,对某时间产生的背景、发展因果、逻辑关系或构成因素、基础水平和效益功能等进行深入分析,

说明其先进性和适用性,从而辨清优劣,达到选择的目的。

3) 核查法

通过对有关信息所涉及的问题进行审核和查对,来优化信息的质量。这可以从以下三方面入手:一是核对有关原始材料或主要论据,检查有无断章取义或曲解原意等情况;二是按该信息所述方法、程序进行可重复性检验;三是深入实际对有关问题进行调查核实。

4) 引用摘录法

信息的相互引用摘录表明了各信息单元之间的相互关系,一般来说,被引次数较多或被本学科专业权威出版物引用过的信息质量较高;被文摘索引等著名检索工具摘录或在综述评论文章中有所反映的信息,其价值也比较大。

5) 专家评估法

对于某些内容专深且又不易找到佐证材料的信息,可以请有关专家学者运用指标评分法、德尔斐法、技术经济评估法等方法进行评价,以估测其水平价值,判断其可靠性、先进性和适用性。这类方法准确度高,但费用较大,一般只用于选择那些十分重要的信息成果。

2. 确定标识

经过优化选择的信息要进行加工整理,确定每条信息在社会信息流这一时空隧道中的“坐标”,以便使人们在需要时能够通过各种方便的形式查寻、识别并获取该信息。要想在四维信息空间中标定信息的具体方位,达到容易提取的目的,关键是要确定该信息所具有的区别于其他信息的基本特征,并以适当的形式描述,使其成为该信息的标识。一条信息之所以有别于其他信息,主要是因为它与其他信息的外表和内容两方面的特征都有所不同。信息外在特征包括名称、类型、表现形式、生产者、产地、日期和编号等;信息内容特征是指该条信息所涉及的中心议题和学科属性等。

1) 著录项目的确定

按照一定的标准,对存在于一定物理载体的信息的外表特征和物质形态进行分析、描述加工和记录的过程称为著录。著录项目中包括信息的题名、作者、出版、主题、学科和号码等各个著录事项。著录项目依描述对象及其加工要求而异。任何一个著录项目都有可能成为未来数据库的检索入口。因此,著录项目的选取恰当与否,不仅关系到能否准确地代表所描述的信息,而且影响到数据库的功能和检索效果。著录项目的选择要遵循完整性、标准化、方便性、低冗余性和灵活性等原则。

2) 信息外在特征的加工

在著录的基础上,若干著录项目按照一定的逻辑以一定的格式组成款目,众多款目再依一定的规则排列,即成为信息加工的最终产品——目录、题录、文摘索引或数据库等。

3) 信息内容特征的加工

信息内容特征的加工是指在对信息内容进行分析的基础上,根据一定规则给信息的内容属性以标识,并做出描述的过程。这一过程通常称为信息标引。根据标引过程中所给出的标识形态和性质的不同,信息标引通常可分为以学科分类代码为基础的分类标引和以主题语词符号为基础的主题标引。

① 分类法。分类法是按信息内容的学科属性来系统提示和组织信息的方法。通过分类标引,可以将具有共同学科属性的信息类聚在一起,并依据各类信息之间的学科关系把所有信息组织成一个有层次、有条理的整体。

② 主题法。主题法是按信息内容的主题名称来系统提示和组织信息的方法,它是一种以规范化或未经规范化的自然语言作为信息主题标识的方法。所谓主题,是指某条信息所论及或涉及的事物。表达主题的语词称为主题标识(主题词)。通过主题标引,可以把有关同一主题的信息集中在一起,并将其按字顺排列起来。

3. 组织排序

对每条信息的各种内外特征进行描述并确定其标识之后,必须按一定规则和方法把所有信息记录组织并排列成一个有序的整体,才能为人们获取所需信息提供方便。

1) 分类组织法

分类组织法是依照类别特征组织排列信息概念、信息记录和信息实体的方法。它符合人类的认知习惯,是一种普遍使用的信息组织方法,在社会活动的各个领域,均可找到大量的实例,如分类目录、分类索引、分类词典、分类广告、分类展品陈列和分类统计报表,等等。对信息实施分类组织,需要对每一个组织排列对象的类别特征进行分析,为它们赋予分类代码或其他形式的类别标识,然后再按照类别的不同或分类代码的次序排列起来。

2) 主题组织法

主题组织法是按照信息概念、信息记录和信息实体的主题特征来组织排列信息的方法。该法给人们提供了一种直接面向具体对象、事实或概念的信息查寻途径。学术论文以及书刊内容的组织中采用的标题、章节次序等可视为较简单的主题组织法。

3) 字顺组织法

字顺组织法是按照提示信息概念、信息记录和信息实体有关特征所使用的语词符号的音序或形序来组织排列信息的方法。这是一种完全采用语词符号的发音和结构特征作为排序依据的方法,故而操作简单,应用广泛。各种字典、词典、名录、题名目录等大多采用字顺组织法。但是,用这种方法组织信息概念时,排序结果只能显示表达信息概念的语词符号在音、形方面的联系和差异,很少或基本上不能反映信息内容之间的联系。

4) 号码组织法

号码组织法是按照每条信息被赋予的号码次序或大小顺序排列的方法。某些特殊类型的信息,如科技报告、标准文献和专利说明书等,在生产发布时都编有一定的号码。对于其他信息产品,有时为标明其来源、类型、性质和生产日期等,同样也需要给予相应的编号和代码,按号码对信息进行组织排列十分简便易行,尤其适用于计算机信息处理、存储和检索。

5) 时空组织法

时空组织法是按照信息概念、信息记录和信息实体产生、存在的时间、空间特征或其内容所涉及的时间、空间特征来组织排列信息的方法。任何事情都是在特定的时间与空间中产生、存在和运动着的。因此,时空组织法可用于对任何信息概念、信息记录和信息实体的组织排序。其结果,或者是按时间顺序把有关信息排列成一定的次序,如年鉴、大事记、历史年表等;或者是按空间位置把相关信息组织在一起,如国家、地区、城市和乡镇等;或者是交替运用时空特征以形成多层次的信息集合,如地方志等。

6) 超文本组织法

超文本(hypertext)组织法是一种非线性的信息组织方法,它的基本结构由节点(node)和链(link)组成。节点用于存储各种信息,链则用于表示各节点(即各知识单元)之间的关联。利用计算机处理技术,把文本信息中若干可产生联想的内容以显而易见的线性的方式组合在一起,即通过建立各节点间的超文本链接(hypertext link),构成相关信息的语义网络,就可以实现超文本的信息组织方式。

4. 改编重组

信息改编和重组工作是指根据用户需要将分散的信息汇集起来进行深层次加工处理,提取有关信息并适当改编和重新组合,形成各种精约化的优质信息产品。按照加工深度的不同,信息改编和重组的方法主要有汇编法、摘要法和综述法三种。

1) 汇编法

汇编是选取原始信息中的篇章、事实或数据等进行有机排列而形成的,如简报资料、文献选编、年鉴名录、数据手册和音响剪辑等。运用汇编法,基本上不需要对信息内容进行复杂的分析和浓缩,只要抽取有关的信息片断按一定方法编排加工,就可以方便及时地汇集某一专题或专业的资料。

2) 摘要法

摘要是对原始信息内容的浓缩加工,即摘取其中的主要事实和数据而形成的二次信息产品。因其所摘内容大多来自用文字记录下来的信息,故又称文摘。按加工目的,可分为报道性文摘、指示性文摘和报道/指示性文摘。

3) 综述法

综述是对某一课题某一时期内的大量有关资料进行分析、归纳和综合而成的具有高度浓缩性、简明性和研究性的信息产品。按综述编写手法的不同,可分为叙述性综述和评论性综述。叙述性综述只就有关某一专题的事实、观点和数据等大量资料客观、全面地综合叙述,不加评论地综述作者意见,不发表自我见解;评论性综述又称述评,它是在叙述性综述的基础上,加入综述作者对有关问题的见解和评论而形成的比较复杂的综述。

6.2 知识组织概述

知识组织是信息组织的高级发展阶段,其在概念的内涵、组织原则和组织方式等方面均与信息组织有所不同。

6.2.1 知识组织的概念

所谓知识组织,是指对知识客体进行诸如整理、加工、揭示和控制等一系列的序化、系统化的活动。知识组织的本质是对知识及知识间的关联进行揭示和组织。组成知识的基本单元是概念,概念间存在着各种逻辑关系。知识组织就是通过一定的技术手段,将概念及其关联关系组织成概念系统的一系列活动^[2]。

知识组织与信息组织在组织原理、组织深度和组织结果方面均有所不同。信息组织是对处于自然状态的信息在时空序列上的重新排列,而知识组织是一种深入到人类认知过程的探讨;信息组织只在语义信息组织时才涉及内容的变化,而知识组织以语义内容为基础;信息组织的结果是有序的信息,知识组织的结果是新产生的附加知识或浓缩知识。

也就是说知识组织的重点在于对知识的关联进行有效的描述、处理和表示。它以知识结构为描述对象,这是知识组织有别于信息组织的典型特征。知识结构不是线性的、等级式的,而是呈网状的,概念是知识结构的基础要素;知识结构的表现形式就是具有各种关系的概念群;知识组织就是要将知识结构中的概念关系揭示出来。

知识组织的分类如下所示。

(1) 从知识的形态看,可分为主观知识的组织和客观知识的组织。主观知识的组织在人的大脑中进行,表现为复杂的神经生理活动,人工智能和认知心理学等;客观知识的组织是通过人的认知进行分类,并凭借一定的方法完成的。本书中的知识组织主要指后者。

(2) 从知识运动的流程看,可分为生产过程中的知识组织、知识交流过程中的知识组织以及知识利用过程中的知识组织。

(3) 从知识的表述程度看,可分为显性知识组织和隐性知识组织。

(4) 从知识组织的主动性看,可分为知识被组织和知识自组织,知识自组织是指人工智能领域的知识组织或没有统一管理的大规模的知识组织活动中所达到的一种平衡状态。

6.2.2 知识组织的原则

知识组织活动必须同时面向两个方面:一是面向知识客体本身,因为知识客体是所有知识组织活动的直接本体对象;二是面向用户,即必须按用户需求去组织知识^[3]。

1. 全面性原则

知识组织的全面性是指知识组织系统在知识组织过程中所能占有或连接的知识客体——知识资源的范围及其数量的全面程度。知识资源占有或连接知识资源的全面性,是任何一个知识组织系统都竭力追求的目标,因为这是提供知识服务的基本物质基础。从一定意义上说,知识组织系统占有或连接知识资源的全面性程度是衡量系统服务能力的基本标准。

2. 客观性原则

客观性原则主要体现在知识表示或知识重组环节上。它是指在知识表示或知识重组的过程中,对知识客体的形式特征和内容特征必须如实客观地加以描述或揭示,而不能添加任何主观的推断或评判。这一原则的目的是为了反映知识客体的原貌,尊重知识原创者的意愿。

3. 充分性原则

充分性原则主要体现在知识表示环节上。它是指在知识表示过程中尽可能全面深入地显示知识客体的内容和形式信息。充分性原则体现在广度和深度两个方面。充分性原则的广度,是指多角度、多途径、全方位地表示知识客体的内容和形式信息。充分性原则的深度,是指揭示知识客体内容信息的深入程度。

4. 有序性原则

有序性原则是指经过组织后的知识客体必须呈现出有序状态。知识序化是知识组织的首要任务。狭义的知识组织就是指知识序化。知识序化包括载体序化和环境序化两方面。

5. 标准化原则

标准化是知识组织的生命。没有标准化就不可能形成知识组织的方法体系。知识组织的标准化原则,就是指知识组织的每一种方法及其操作,都要按一定的标准进行。知

识组织标准化原则的内容包括科学性原则、简化原则、统一原则、协调原则和稳定继承原则。

6. 用户可近性原则

把知识组织成什么样子？这一问题不仅取决于知识本身的特性，还要取决于用户的需求，因为用户是知识组织成果的利用者和评判者。“可近性”（accessibility）概念集中反映了用户对知识组织系统的要求。它是指用户与知识组织系统之间的相互关系在用户意识中的反映。其内涵包括三方面：① 用户与知识组织系统之间的物理可近性，这里主要指两者之间的时空关系；② 用户与知识组织系统之间的智力可近性，即两者之间在知识结构上的匹配关系；③ 用户与知识组织系统之间的心理可近性，即两者之间在情感、态度、感觉等心理反映上的亲疏关系。

7. 经济性原则

知识组织的经济性原则，是指在知识组织过程中，既要考虑知识组织的理想目标状态，又要考虑系统和用户的经济承受能力。开展知识组织活动，必然要采取一系列的技术手段，而任何技术的使用都必须消耗一定的人力、物力和财力。技术与经济的这种相互依赖和相互统一的关系，使任何技术问题都与经济问题联系在一起。这就迫使知识组织系统在设计技术手段时，不得不重视技术本身的经济性能问题。

8. 逻辑性原则

逻辑性原则主要体现在知识组织方法体系的建构过程中。这里的“逻辑性”，包含层次性、联系性、体系性、规则性和规范性等含义。知识组织的逻辑性原则，一般表现在思维逻辑、概念逻辑和方法逻辑等方面。

9. 思想性原则

每一种知识组织方法，都是人们在实践基础上经过思维总结而产生的。而人的思维活动无一不受世界观和方法论的影响。所以，从严格意义上说，每一种知识组织方法都体现出或明显或隐含的思想性特征。这里的“思想性”包含有政治倾向性、伦理道德主张、思维倾向性和科学观念等含义。

10. 发展性原则

毋庸置疑，任何知识组织方法都不是一成不变的。任何知识组织方法，都要不断更新和完善自身的结构和功能，以适应和满足社会和科学技术发展的需要。知识组织的发展性原则是指知识组织系统在尊重历史和实践的基础上，重视用新思想、新技术不断完善自身的结构和功能，以适应和满足社会和科学技术发展的需要。知识组织的发展性原则，主要体现在“以旧出新”和“以新代旧”两个方面。在“以旧出新”方面，主要表

现为原有知识组织方法的不断修缮过程；在“以新代旧”方面，主要表现为用新方法代替旧方法。

6.2.3 知识组织的方式

知识组织的方式大体上可以分成两种：一种是以知识单元为基础的知识组织方式，另外一种是以知识关联为基础的知识组织方式^[4]。

(1) 以知识单元为基础的知识组织方式是以知识单元为处理对象。知识单元是经过专家精心评价、筛选、提取和测试之后获得的“浓缩”的知识。以知识单元为基础进行知识组织就是将知识单元或知识单元集合中的知识因子抽出，对其进行形式上的组织。由于只是对知识因子进行组织，而并未改变因子间的联系，因此在此过程中没有产生新的知识。为了提高工作效率，节约用户时间，就要向用户提供“纯信息”，即向用户提供其所需的知识和信息。由于这些知识信息都是事先经过精心选择、评价后存入的知识单元，因此用户无须再查原始文献。这种方式是当前知识组织的主流和方向。

以知识单元为基础进行知识组织，首先要确定知识基本单位，然后存储这些基本单位并进行有序排列，检索则通过模式匹配，着眼于查找已存储的知识基本单元。整个系统只关心用户需要什么知识，而不关心需要这些知识的目的。由于知识基本单元的不确定性和模糊性，这种知识组织方式难以满足人们对知识的多样化需求。

(2) 以知识关联为基础的知识组织方式是在相关领域中提取大量知识因子，并对其进行分析和综合，形成新的知识关联，从而产生出更高层次上的综合的知识产品。由于改变了知识因子间的原有联系，因此其结果既可以提供新知识，也可以提供关于原知识的评价性或解释性知识。

这种组织方式不仅是知识单元的增加，而是知识在更高层次上的网络化和综合化，因而是内容的提纯和浓缩。

以知识关联为基础的知识组织方式，最突出的特征就是具有联想、判断、比较、推理分析和学习等新兴的知识处理和组织功能，它是人类关于知识的研究与现代信息技术相结合的产物。

6.3 知识组织的方法

为克服传统的信息组织方法的不足，如不能构建领域知识的整体结构，揭示概念间关系方面能力有限，不能将知识可视化，用户无法直接获得与知识相关的资源等，新的知识组织方法和工具应运而生。这些新的知识组织方法以概念为基础，注重概念间关联关系的组织，在网络技术的支持下，能以可见的方式揭示和展示知识结构，其用途也日

趋多元化,除了支持信息和知识检索,还可用于回答问题,进行机器翻译等基于推理的各种用途。

知识组织的方法类型多样,比如,依据知识内部结构特征,可分为知识因子的组织和知识关联的组织方法;依据知识的不同存在形态,可分为主观知识的组织方法和客观知识的组织方法;依据组织的语言学原理,可分为语法组织方法、语义组织方法和语用组织方法等。本章针对知识结构中概念关联体系的构建问题,重点介绍语义词典、概念图、主题图、本体、语义网、特定领域分类法(Taxonomy)、网络分类法和分众分类法(Folksonomy)八种知识组织方法。

6.3.1 语义词典

对概念进行语义分析,语义词典是不可或缺的基础性资源。美国Princeton大学的WordNet^[5]就是一个非常好的范例。它是由Princeton大学的心理学家、语言学家和计算机工程师联合设计的一种基于认知语言学的英语语义词典。作为心理语言学家的实验品,WordNet的最初设计并不是受到计算语言学的影响而是直接为自然语言处理服务的。20世纪80年代末,由于语义计算的需要,计算语言学家发现了WordNet并将之应用于自然语言处理中涉及语义分析的诸多领域。WordNet的基本思想简单明确,它的形式化做得很彻底。目前,WordNet已经成为一个事实上的国际标准,从EuroWordNet发展的事实不难看出,WordNet框架的合理性已被词汇语义学界和计算词典学界所公认。

WordNet是一个覆盖范围广泛的英语词汇语义网,它的独特之处在于它是依据词义而不是依据词形来组织词汇信息的。WordNet使用同义词集合(synset)代表概念(concept),词汇关系在词语之间体现,语义关系在概念之间体现。WordNet构造的核心是如何表示词汇概念节点,以及在这些概念节点之间建立起各种语义关系。WordNet将名词、动词、形容词和副词各自组织成一个同义词集合,每个同义词集合都代表一个基本的语义概念;同时在概念间建立不同的指针,以表达上下位、同义和反义等不同的语义关系。这样就构成了一个比较完整的词汇语义网络系统。经过这样的过程,原本抽象的概念就被形式化了,变得具体而且可以通过词汇意义加以操作,概念之间还可以建立多种语义关系的联系和推理。这是在传统的义素分析法外简单而有效地表达词汇语义的另一种新的方式和途径。

WordNet中词汇概念的语义关系主要包括:上下位、同义、反义、整体和部分、蕴涵、属性和致使等。WordNet Version 2.0中包含了42 000多条词形相关名词与动词之间新的链接,描写了四类实词10 000多个概念节点和超过5 000 000个的语义关系,形成了一张庞大的概念语义网络。

6.3.2 概念图

概念图（concept map）是利用概念以及概念之间的关系表示和组织结构化知识的一种可视化方法，于 20 世纪 60 年代由美国康奈尔大学 Joseph D. Novak 教授等人提出，主要应用于教育学方面。将概念图引入到知识组织领域，一方面是为了让每个人都可以根据自己的知识结构组织知识，而从所有人的知识结构中可以抽取共享的知识结构；另一方面，概念图的构建过程也是个人隐性知识显性化的过程，存储人们所构建的概念图，即完成了隐性知识的沉淀。

概念图用节点表示概念，用连接线和连接词表示概念之间的关联。概念和一些连接都是有标签的。连接可以是无向的、单向的或者双向的。概念图具有三个要素：① 命题（propositions），指由两个以上的概念及其关系构成表达意义的陈述；② 层级结构（hierarchical frameworks），指概念图中的概念按照宽泛概念在上、具体概念在下的顺序排列形成层级结构；③ 交叉连接（crosslinks），指不同分支中的概念之间形成的连接关系。概念图将相关概念置于圆圈或者方框之中，然后用连线将相关的概念和命题连接，连线上标明两个概念之间的意义关系，如图 6.1 所示。

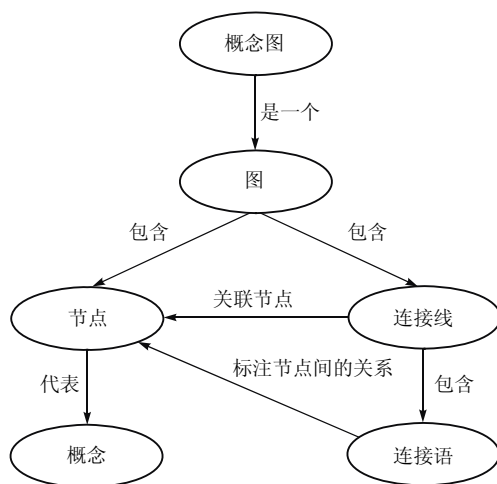


图 6.1 概念图示例

根据Novak教授提出的概念图构建准则，概念图的构建包括以下四个步骤^[6]：

（1）概念选取，指列出关于某个主题的所有重要概念，确定主题或核心问题，确定将要利用概念图进行阐述的中心主题，并找出与该中心主题相关的一系列重要概念。这些提出的概念是回答中心主题时需要用到的，要将其罗列出来。

（2）概念分类，分为广度结构和深度结构。广度结构根据概念间关联性强弱将概念划分为不同的分支；深度结构将不同分支中的概念按照概念的宽窄由上至下排列。将列

出来的概念进行排序,是将最抽象、最概括的概念放在最顶端,其余概念按层级从上到下依次排放。这一步骤使包含关系清晰化。

(3) 定位中心概念,然后通过连接线连接概念。首先,获取中心主题概念,作为整个概念图的根节点。然后,从中心节点出发连接每个分支的中心节点,形成概念图的第二层,并注明连接词。继续上述过程,直至连接完所有概念。连接线的作用是将相关概念连接起来,并在连接线上标明连接语,连接语要清晰地表达两个概念之间的关系,使之成为简单、有效的命题。当大量相关的概念连接后会形成有层次的、对应知识、命题及中心主题的知识架构。

(4) 连接交叉概念。仔细研究概念图中的各个概念,看它们是否存在交叉关系。若存在,用交叉连接将处于概念图不同区域或子区域中的概念连在一起,并注明连接语,其目的是详细阐述不同领域概念的关联关系。

概念图的构建一般是利用概念图软件工具由人工构建。它也可由多个用户协作完成。概念图的协作创建可以采用多种形式,既可以面对面地进行,也可以远程进行,既可以同步,也可以异步。

基于概念图的知识组织方式与传统信息组织方式有着很大不同,其优势表现在如下十个方面。

1) 基于概念图的组织方式保证了领域知识的整体性

概念图采用半等级结构的方式来组织知识及资源,既遵循等级结构又允许交叉连接。它采用将知识领域作为一个整体来建立和维护领域的知识组织结构。这种非线性、半等级的组织方式更加灵活,不仅能反映某单一知识领域的知识结构,同时也能反映不同领域知识之间的内在关系,从而体现知识之间以及学科领域之间的不可分割性,形成系统完整的知识结构。使用户能够从整体上来把握领域知识结构,保证了学科知识体系的系统性及完整性。

2) 概念图的知识表征是基于语义的

与传统的信息组织方式不同,基于概念图的知识组织是建立在知识之间以及知识与相关媒体资源之间关系基础之上的。概念图利用严格的语义描述机制来揭示各种关联关系,因此它实现了基于语义的知识表征与知识组织。

3) 概念图实现了领域知识可视化

概念图不仅能够深层次地揭示概念间及概念与资源间复杂的内部关系,而且它利用图形的形式直观地描述领域知识架构及相关媒体资源,从而将这些关系可视化。概念图是领域知识结构的可视化方式之一。

4) 使知识及资源更为集中

概念图弥补了传统信息组织导致信息分散的缺憾。概念图以图形方式揭示知识之间

以及知识与资源之间的关系,并将知识结构与相关媒体资源结合起来,形成资源的集合——概念图知识模型。这种集合式的组织方式可以有效地增强信息资源的可移植性,并大大提高资源的可获得性及共享性。

5) 方便用户对相关媒体资源的获取

概念图工具提供了一套将概念和与概念相关的媒体资源集合起来的机制,并将其组织成为概念图知识模型。它支持各种规模的概念图知识模型的开发,而且不限制资源的位置和概念图的地理归属。用户可以直接通过知识导航或检索工具很容易地从概念图知识模型中获得所有所需的、与概念相关的各种媒体资源及内容。

6) 概念图的构建相对简单,支持同步及异步协作

单一概念图的构建过程十分简单,构建大型的复杂领域概念图知识模型可以采用与其他用户协作的方式来完成。基于概念图的知识组织方式提供了同步协作和异步协作机制来支持用户协作构建知识模型。在概念图构建过程中,网络上来自任何地方的概念图用户可以进行知识模型构建的同步或异步协作及共享。也就是说,不同地点的多个创作者可以同时或先后构建及修改同一概念图。这种方式可以综合多个创作者的思维,最终将不同见解统一起来,形成功能更加强大、资源更加丰富的概念图知识模型。而且用户还可以把他们的知识模型发布到遍布互联网的概念图服务器上,或者对其他用户的知识模型进行共享、协作和探索。

7) 基于概念图的导航功能更强大

概念图作为一个导航于大型知识域的浏览器是特别有效的。概念图揭示了概念间以及概念与资源间的关系,将知识组织上升为语义的高度,因此其导航也是基于语义的导航。概念图知识模型为目标资源提供了标识,其目的是要让用户知道,下面的链接包含了领域或上下文的变化,这样用户就意识到它来自不同的域,并可以在没打开概念图之前做出选择。概念图较好的导航功能可以提高用户的使用效率。

8) 概念图的组织方式提供了更好的学习功能

概念图这种知识组织方式在学习者学习过程中具有独特的优势。在视觉角度方面,概念图以可视化形式进行知识表征和表现复杂的领域知识,用图形呈现知识之间的联系,使该领域的知识更加整体化、系统化和直观化。在知识表示方面,概念图能够构造一个清晰的知识网络,便于用户对整个知识架构的整体掌握,有利于直觉思维的形成,促进知识的创新。在抽象角度方面,概念图不但可以提高用户综合利用知识的能力,增进用户对知识的整体理解,而且可以改变用户的知识结构,提高其学习效率,提升用户的学习技能。

9) 容易扩展

概念图具有灵活的可扩展性特点,一幅概念图可大可小,设计也不固定。用户可以

根据自己对专业领域知识理解的不断深入，随时对已经构建的概念图进行修改及完善。

10) 方便知识及资源的管理和维护

概念图采用文件夹形式管理知识及资源，而且知识模型是集合式的，这就大大方便了用户对概念图的维护。

6.3.3 主题图

主题图（TM，Topic Map）是一种新兴的知识组织和知识表示工具，是以图形方式来展示某一资源库的知识结构的。它是一种用于描述信息资源知识结构的数据格式，可以用来定义某一主题概念所在的资源位置，也可以用来表示主题概念之间的相互联系。主题图实际上在信息资源的上层构建了一个结构化的语义网，它独立于技术平台，描述主题之间的关系及主题与具体资源的联系，通过揭示主题概念之间的关系，将用户指引到相关的资源。

主题图与前面介绍的概念图在以图的方式展现知识结构方面，非常相似。但是，主题图不仅仅是知识组织工具，还是一种知识表现工具。其表现方式除了以直观的图的方式展现外，还可以提供以被机器理解和处理为目标的标记语言标识的文件方式。主题图还将概念结构与具体相关资源连接起来。

1. 主题图的结构

主题图的构成包括主题（topic）、关联（association）、事件（occurrence）三个主要要素；以及知识层（knowledge level）、信息/资源层（information/resource level）两个层次。图 6.2 所示为主题图的基本结构。

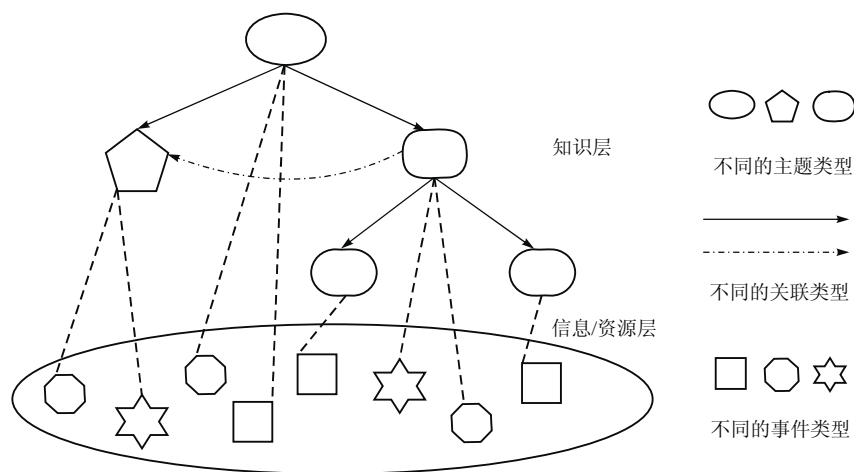


图 6.2 主题图基本结构

下面详细介绍主题图的基本组成要素。

主题是主题图中最基本的概念。它可以是人们能够描述出来的任何事情，如具体的人或物，或抽象的概念，等等。通常意义上，主题图中的一个主题代表现实世界中的一个东西，这里称这个“东西”为一个“话题”(subject)。主题与话题之间是一一对应的，也就是说，一个主题代表一个话题，同时一个话题也仅能被一个主题所表示。根据主题的特点和所代表“话题”的不同，主题还可以被分为许多个主题类型，如图 6.2 所示，知识层中的不同图形代表不同的主题类型。主题类型可以作为一项特殊的索引，方便用户快速地浏览整个知识结构。

关联是用来描述主题之间关系的要素，它是在主题图中起语义作用的关键要素。关联通常被双向定义。例如，“出版”关联的逆关联就是“被出版”，正向关联和逆向关联同时存在于两个主题之间。但是，在有些时候，关联不是对称的，而是单方向的。例如，“吃胡萝卜一定能促进身体健康”，并不意味着“身体健康一定要吃胡萝卜”。另外，大部分时候，关联是可以传递的。例如，“A 包含 B，B 包含 C，所以 A 包含 C”。但是有些时候关联是不可以传递的。例如，“A 教 B 数学，B 教 C 数学，并不意味着 A 教 C 数学”。上述属性可以用下面的式 (6.1) 和式 (6.2) 加以描述

$$\forall A \rightarrow B, \text{不是总有 } B \rightarrow A \quad (6.1)$$

$$\forall A \rightarrow B, B \rightarrow C, \text{不是总有 } A \rightarrow C \quad (6.2)$$

此外，与主题类似，关联也可以被分为许多关联类型。如图 6.2 所示，不同的有方向的线代表不同的关联类型，它也可以作为一项索引指导用户对整个知识结构的浏览。

事件可以是与主题相关的一切信息资源，它以任意的形式存在，如文章、视频和图表，等等。事件通常被存储在信息/资源层中。一个主题可以连接多个事件，一个事件也可以关联多个主题。与主题和关联类似，事件也可以分为许多类别，不同的类别可以表示事件不同的存储形式。如图 6.2 所示，下方椭圆形中的不同形状代表不同的事件类型，它们共同构成了主题图的信息/资源层。

主题图还包含其他元素用以细化上述三个基本要素和描述信息资源。关联角色 (association roles) 用来说明发生关系的主题概念的角色，比如 A 是 person，B 是 place；关联类型 (association types) 用来说明发生的具体关系；关联角色类型 (association role types) 用来说明关系的类型。比如，A 与 B 之间发生的关系属于 action。事件角色 (occurrence roles) 用来说明揭示的信息资源的类型，范围 (scope) 元素用来限定主题概念的定义范畴，侧面 (facets) 则用来描述信息资源的其他特征。

属种关系是概念之间非常重要的一种关系，但是仅有属种关系，却不能完全反映概念之间的所有联系。在主题图中所定义的概念之间的联系，既包括属种关系，还可包括其他关系，如因果关系、危害关系、平行关系、大于关系，等等。而且，在主题图中通过定义范围 (scope) 可从不同的角度展现概念之间的属种关系。

2. 主题图的构建方法

2003年, Empolis GmbH出版的《White Paper—The Topic Maps Handbook》一书中详细介绍了主题图的构建过程^[7]。

主题图的构建主要分为以下六个阶段。

1) 目标和定义阶段

在目标和定义阶段中, 主题图开发者要完成以下工作:

- (1) 定义主题图应用的领域, 即主题图要描述的对象和要包含的内容;
- (2) 定义功能上的要求, 即主题图的用户及使用该主题图的目的;
- (3) 定义主题图的框架, 即确定主题图应该包含的“话题”, “话题”之间的关联方式, 以及构建一个有效且一致的主题图的方法;
- (4) 选择实现主题图的工具, 即实现主题图的软件和技术;
- (5) 选择组装整个主题图的方法, 即选择人工或自动的方法生成事件信息, 确保构建的主题图满足用户的初始需求, 检查该框架是否需要改进;
- (6) 选择维护主题图及其应用的方法, 即确保主题图可以长期的、与时俱进的运行下去的方法。

2) 分析阶段

分析阶段确保了主题图应用的结果, 保证了主题图解决问题的一致性。在这个阶段中, 主题图开发者需要完成以下工作:

- (1) 列出该主题图所要完成的全部任务及各个任务的优先顺序; 考虑用户的实际需求、功能要求、设计要求、系统要求和效果要求, 以及预算、时间和人力资源成本, 等等;
- (2) 列出可以用到的全部资源, 包括已经存在的知识结构框架, 如本体、主题分类、分类法、关系框架、OO 框架、元数据词表和语料库, 等等;
- (3) 利用头脑风暴法, 列出与该领域相关的全部“话题”; 在此基础上, 形成主题并进行分类。

3) 设计阶段

在分析阶段的头脑风暴过程中, 已经初步定义了一些主题类别和关联角色, 设计阶段将进一步明确地定义并组织这些信息资源。在这个阶段中, 主题图开发者要完成以下工作:

- (1) 区分资源、主题和主题类别;
- (2) 从有效的资源中识别出事件和事件类型, 并从用户的角度加以定义;
- (3) 观察主题类别, 识别出有意义的关联类型;
- (4) 确定主题类别的层次, 区分“is-a”关联和“is-a-kind-of”关联;
- (5) 合并意思相同、表达方式不同的主题。

4) 生成阶段

以上三个阶段是主题图构建的纸上谈兵的过程,而生成阶段是主题图真正的构建阶段。在这个阶段中,主题图开发者需要利用以下三种技术中的一种:

- (1) 利用一个 ASCII 编辑器或 XML/SGML 编辑器编写 XTM (或 HyTM) 代码;
- (2) 编写将主题图自动转化为 XTM 的伪代码;
- (3) 利用一个主题图软件,如 TM4J、Protege 2000,等等。

5) 测试阶段

对于主题图应用效果来说,该阶段是一个关键的阶段。在这个阶段中,主题图开发者和用户需要分别对已有的主题图进行测试。

开发者应该考虑如下问题:

- (1) 是否所有的元素都有用?
- (2) 是否有什么元素不见了?或是否有一些元素放在了错误的地方?
- (3) 各元素划分的粒度是否正好?
- (4) 是否所有的名称都一致?
- (5) 查看那些只有一个子类或有許多子类的类别,考虑模型建得是否正确?

开发者进行测试之后,开始用户测试阶段。此时,用户应该考虑如下问题:

- (1) 所构建的主题图是否友好?
- (2) 是否要求的元素都在其内?
- (3) 用户能否找到想找的东西?

在这两个阶段之后,开发者根据测试结果重新调整主题图的相应内容,然后进行以下三个测试:

- (1) 一致性测试,即测试全部要求是否满足;
- (2) 统计测试,即统计主题、关联和事件的数量,以及它们在各类别中的数量;统计每个主题所发挥的关联角色的情况;
- (3) 功能测试,即测试全部功能是否满足要求。

6) 文件化阶段

文件化阶段可以帮助主题图使用者更好的使用、维护主题图。文件化的结果可以作为主题图的一部分,内嵌到主题图中。它可以以事件类型的方式出现在主题图的信息/资源层中。例如,将帮助文档作为一个类型为“帮助”的事件安置在信息层中。

3. 主题图的优势和缺陷

主题图有两个层次,其中主题和关联构成主题图的知识层,事件构成主题图的信息/资源层,两层之间相互独立,这一特性使得主题图具有如下优点:

- (1) 主题图的知识层可以移植到任意信息资源上,对于表达相同知识结构的问题来

说,这种特性避免了重复构建领域主题图的麻烦,避免了重复劳动带来的效率和劳动成本问题。

(2) 不同的主题图的知识层可以覆盖到同一个信息/资源层上,换言之,对于同样的信息资源,不同的主题图可以从不同的角度来描述同一个领域的问题。

(3) 当知识层中的知识结构发生变化时,可以直接对知识层进行修改,避免因修改数据库而造成不必要的麻烦。

但主题图也有一些缺陷,如在主题图中没有定义类的继承,没有明确定义推导出内涵的知识的关联属性和推理规则,缺少用于有效性检查的一致性约束等。

6.3.4 本体

关于本体(ontology)的相关概念在第5.3.7节中已经做过介绍。而将本体的思想引入到知识组织领域中,其根本动因是解决不同系统、不同用户之间的认知差异,为人與人之间、人与计算机之间以及计算机系统之间建立共享的概念框架,实现知识共享以及互操作。

正是由于本体能够实现对知识结构的描述和揭示,具有知识组织的功能,从而成为知识组织的主流技术之一。

构建一个本体,可用来解决:① 在用户间或软件代理间达成对于信息组织结构的理解和认识;② 可以复用专业领域知识;③ 使专业领域内的假设变得更加明确;④ 将专业领域的知识从运筹学、知识管理的环境中剥离出来;⑤ 分析专业领域的知识等。

构建本体的基本原则如下所述。

(1) 明确性和客观性,指所构建的本体应该用自然语言对术语给出明确、客观的语义定义。

(2) 完整性,指所给出的定义是完整的,能表达特定术语的含义。

(3) 一致性,指知识推理产生的结论与术语本身的含义不会产生矛盾。

(4) 最大单向可扩展性,指向本体中添加通用或专用的术语时,通常不需要修改已有的内容。

(5) 最少约束,指对待建模对象应该尽可能少地列出限定约束条件。

下面介绍几种常用的本体构建方法。

1. TOVE 法

TOVE 法,又称 Gruninger&Fox 评价法。TOVE 是指多伦多虚拟企业(Toronto Virtual Enterprise),专用于构建 TOVE 本体,由多伦多大学企业集成实验室研制。TOVE 本体包括企业设计本体、工程本体、计划本体和服务本体。TOVE 流程如图 6.3 所示。

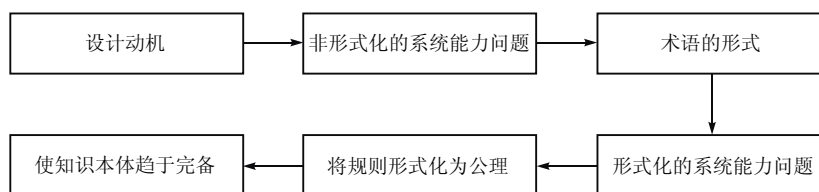


图 6.3 TOVE 流程

2. 骨架法

骨架法(skeletal methodology), 又称 Enterprise 法, 专门用来构建企业本体(enterprise ontology), 是有关企业建模过程的本体, 它由爱丁堡大学人工智能研究所发起开发。

骨架法构建本体的基本步骤包括:

- (1) 确定本体应用的目的和范围;
- (2) 本体分析, 即定义本体内所有术语的意义及术语之间的关系(即定义“类”和“关系”);
- (3) 本体表示, 即用相应的表示语言(又称置标语言)来表示本体;
- (4) 本体的评价, 即评价本体质量;
- (5) 本体的建立, 即对所有本体按(4)中的标准进行检验, 符合要求的即可保存于本体库中。

3. METHONTOLOGY 方法

METHONTOLOGY 方法是由 Polytechnic 大学的研究人员提出的, 其目的是在知识层上构造本体。该方法构建本体的步骤包括:

(1) 建立规格说明书(specification)。这一阶段要产生一份以自然语言编写的非形式化的、半形式化的或者形式化的本体规格说明书。规格说明书至少包括构建本体的目的、实现本体的形式化程度和范围等信息。

(2) 进行知识获取(knowledge acquisition)。知识的来源很多, 可以是专家、书籍、手册、数字、表格, 甚至是其他的本体, 从这些数据源获得知识的关联技术包括头脑风暴法、访谈、文本的形式化或非形式化的分析以及一些知识获取工具。

(3) 概念化(conceptualization)。该步骤是将领域知识组织成概念模型, 用规格说明书中识别的领域词汇表描述问题和解决方案。

(4) 集成(integration)。集成是指重用别的本体中已经建好的定义时, 既可以查看源本体, 选择适合自己概念模型的术语定义, 也可以选择与自己概念模型中的语义和实现一致的术语定义。

(5) 实现(implementation)。实现是指用任何一种形式化的语言编码实现本体。这

需要一套开发环境的支持,其中至少包括词法和语法分析器、翻译器、编辑器、浏览器、搜索器、评价器和自动维护工具。

(6) 评价 (evaluation)。评价是指在本体生命周期的每个阶段和阶段之间,利用某种参考框架对本体、软件环境和文档进行技术判断。评价既包括正确性,也包括有效性。

(7) 文档化 (documentation)。文档化是指在本体建设的每个阶段都应该有对应的文档。

4. IDEF5 法

IDEF5 法是美国 KBSI 公司开发的用于描述和获取企业本体的方法。用 IDEF5 法构建本体有定义课题并组织队伍、收集数据、分析数据、本体初步开发、本体优化和验证五个主要步骤。在构建本体时,确定领域内的核心概念是首要的、中心的任务。为了达到这个目标,必须完成三项工作:一是对领域内的术语进行编目;二是获取用这些术语描述这一领域时的限制条件;三是建立一个模型,当在模型中加入一条特定的描述时,就会产生“适当的”附加声明。

5. 斯坦福七步法

由斯坦福大学医学院开发的七步法,是目前应用最广泛,步骤最详细,应用领域较广的本体构建方法,它的七个步骤如下所述:

第一步,确定本体的专业领域和范畴。

第二步,考查复用现有本体的可能性。如果自己的系统需要与其他应用平台进行互操作,而这个应用平台又与特定的本体或受控词表结合在一起,那么复用现有本体就是最行之有效的方法。

第三步,列出该领域本体的重要术语。具体实施时可以先列出最全的术语表,而不考虑概念间的交叉或重复,然后完善等级体系并定义概念属性,这是构建本体最重要的步骤。

第四步,定义类 (class) 和类的等级体系 (hierarchy)。完善等级体系可以有三种方法:① 自顶向下法,由某一领域中最大的概念开始提取,然后逐渐细化。② 自底向上法,由底层最小的类开始定义,然后将这些细的类组织在一个更加概括的综合的概念之下。③ 综合法,综合以上两种方法。

第五步:定义类的属性,即定义类之间的内在结构。从术语列表中选择一些类后,其余的可作为这些类的后选属性。

第六步:定义属性的分面 (facets)。一个属性由多个分面组成,属性的分面就是属性取值的类型 (value type)、允许的取值 (allowed values) 等有关属性取值的特征。

第七步:创建实例。

源在语义层上的全方位的互联,实现对异构、分布的网络信息和知识的有效检索及访问。

关于语义网络的知识表示,可以参考第5.3.5节。

6.3.6 特定领域分类法

特定领域分类法由 Taxonomy 翻译而来。Taxonomy 是希腊字“taxis”和“nomos”的合成。“taxis”是指分类(classification),而“nomos”是指管理(management)。Taxonomy 是一个很宽泛的词,可应用于任何学科领域。在信息科学领域, Taxonomy 是指对某一特定领域(既包括主题领域也包括机构领域)的信息、知识的分类组织,表现形式为基于概念的树状分类表,它由等级结构和主题词两部分组成,类似于传统分类法和范畴叙词表,但有差异。在知识组织领域, Taxonomy 是指企业信息分类系统(corporate taxonomies, enterprise taxonomies)。

根据 Taxonomy 类表的产生方式, Taxonomy 可以有两种编制方式:一是自上而下。Taxonomy 类表从上往下构建,先选择基本大类、基本结构再定制各个子类目。这种方式需要人工干预,基本大类或基本结构的选择、设计需要根据 Taxonomy 运作的环境而定。二是自下而上。Taxonomy 从下往上构建,先从内容中选择子概念和子类目,然后根据概念之间的关系构建等级关系和整个类目体系。这种方式是基于内容的,可以完全依赖计算机操作完成。两种方式中,由于自下而上方式是完全基于内容的,缺乏人工设计,因此它的弊端是不能将其他实用需求考虑进去,如企业的运作行为、用户的检索习惯等,因而在实际操作中,通常选择自上而下式或者将两种方式结合起来使用。

Taxonomy 在知识组织方面的优势表现在如下方面:

(1) Taxonomy 的等级特性便于用户浏览和发现知识。当企业或组织的内部知识积累到一定数量,新员工对过去的信息和知识就有去“搜寻的需求”。一个 Taxonomy 的展示能帮助用户通过浏览找到知识资源,并且能为用户提供一个友好的方式来发现组织内的知识资源,提高对可用知识的认识,以及提供“意外”发现知识的机会。

(2) Taxonomy 的分级性质提供与语境相关的知识,换句话说,就是用户在浏览时可以对 Taxonomy 中的知识资源进行定位,帮助用户深入了解该组织中的知识结构。此外,对 Taxonomy 中的知识资源进行定位有助于明确主题词表中主题的含义。

(3) Taxonomy 可以用来明确同义词和歧义词的含义,从而提高搜索的召回率和精确率。如果把同义词归纳在一起,用户不管用哪个词汇,都可以找到相关的知识。比如,“大律师”和“大状”都是资深律师的意思,“大状”更口语化些,把这两个同义词归在一起,不管用户查哪个词,同样的结果都会出现。中文里面有很多缩写的名称,比如说“北京大学”和“北大”其实是一样的意思。

(4) Taxonomy 可以被当作特征来过滤搜索结果。当“苹果”分别代表水果、计算机

公司名称、牛仔裤品牌和一部电影名称的时候, Taxonomy 的作用就会充分地展示出来。Taxonomy 的词条可用于缩小搜索结果, 实现更高的精确度。

(5) Taxonomy 可以用来统一不同的内容和语言。在 Taxonomy 的设计过程中, 内容审计能够确保所有现有的内容都被涵盖(如内联网、文件管理系统、博客和网站等)。同时, Taxonomy 还能通过语言等同关系将各种资源与同一主题连接起来。如果某公司的业务范围涵盖全球, 主要语言是中文、英文和其他某一种文字, 那么 Taxonomy 的词汇里面就可以把这三种语言的对应翻译归在同一个地方。

(6) Taxonomy 的设计是以客户为中心的, 并且是直观的。主题词表的设计要经过严格的用户面谈环节, 以更好地发现用户的习惯和行为。其中的很多技术都被用来确定对用户来说最恰当的术语, 如搜索日志审查和卡片分类等技术。

(7) Taxonomy 可以帮助找出知识的差距。通过检查 Taxonomy 或与同类机构的 Taxonomy 进行比较, 可以发现本身的知识差距。例如, 如果 Taxonomy 中缺乏某领域的词条或者在某些节点上词条欠缺, 则表明该组织缺少相应领域中的知识; 如果构建的 Taxonomy 连接很少的内容资源, 也表明组织缺乏某些方面的知识。反过来说, 如果 Taxonomy 与许多领域的资源相连接, 则说明组织内存在某些方面的知识资源优势, 组织可以利用这些知识来取长补短, 据此重新分配人力和资源。

Taxonomy 与本体的关系表现在基本功能方面, Taxonomy 与本体具有一致性, 它们都是概念和概念关系的集合。一般认为, 各种分类法也可以看作是本体, 因此 Taxonomy 本质上也是一种简单化的本体。在体系结构方面, Taxonomy 是单一树状结构, 而本体一般为网状结构, 它们的共同点是两者都针对某一特定领域的知识, 都基于概念, 以等级结构展示主题/术语。在主题/术语选择方面, Taxonomy 和本体都要依赖于领域专家, 具有相对严格的规范, 且通常由少数专家所控制。

6.3.7 网络分类法

网络信息资源具有多元性、交互性和动态性, 网络用户希望能从多种途径和角度来了解信息资源, 而传统的分类法如杜威十进分类法、国际十进分类法、中图法等已不利于充分地揭示和反映多维的知识空间。

针对网络信息资源, 网络分类方法大致可以分为主题分类法、分面组配分类法和学科分类法三种。

1. 主题分类法

主题分类体系采用自然语言表达主题内容, 一个主题对应一个类目。这些类目的选择大多以热点事物主题为类名, 类目不设标识符号, 类目可以细分, 同位类目一般按字

顺排列,但重点类目则可通过类目上置来突出其地位,而不考虑该类目实际所处的层次。最后一级类目通过列表的方式显示与该类目主题相关的超文本链接。主题分类法一般设置 14~18 个一级类目,层次多为 3~4 级。目前大多数网站的分类目录都采用这种分类方法,如新浪、搜狐和雅虎等。由于这种分类方法是依事物分类,能够将相关事物的信息资源组织在一起,因此较易检全同一事物的各方面信息;但缺点是容量小,对网络信息资源的覆盖率较为有限。

2. 分面组配分类法

分面组配分类体系由若干个分面组成,每一个分面的类目可以与其他分面的类目组配,表达专指的概念。分面组配法的原理是:首先确定几个分类标准,即分面;再确定每个分类标准中的若干特征值,即类目;每一个分面的类目与其他分面的类目分别组配,便可形成多个组配类目。由于分面组配分类法专指度高,检索时可以将不同分面的类目进行组配,因此具有较高的查准率。

3. 学科分类法

学科分类法是将网络信息资源按学科属性进行区分和集中的分类法。由于其学科类目与传统的图书分类法不同,所收录的学科类目较为粗放,所以学科分类体系更适于承担以搜集高价值的网络学术信息为主的任务,它的优势在于比主题分类法有更大的容量。

网络分类法的优势如下所述。

(1) 实用性。网络信息分类法大都以事物为中心设置类目,直接以表达事物概念的主题词做类目名称,具有通俗、易懂、简练和透明的特点,符合一般用户的认知水平。

(2) 多维性。为了方便用户从不同的方面、不同的角度检索信息,网络信息分类法大都采用多重列类、重复反映的方式来增加检索途径。例如,一个子类可能隶属于多个母类,一个母类也可能有多个划分标准。

(3) 动态性。由于网络信息处于不断变化之中,新的信息不断出现,旧的信息也会不断地被淘汰,用户的需求也在不断地变化,因此网络信息分类法极具动态性,类目设置是及时更新的,分类表处于不断的调整修改之中,这样就能及时反映新事物和新概念,满足用户的信息需求。有的网站通过设置镜像类目,以链接的方式,在有关事件临近或进行的过程中提前设置类目。事件过后则取消该镜像类目,使分类法具有传统文献分类法所没有的动态性。

但网络分类法也存在很多问题,如基本大类设置不科学,分类标准不明确,类目的划分与排列不当,类目名称不规范,语义含糊不清,不同类目分类层次不均衡,各分类体系之间不统一等。

6.3.8 分众分类法

随着 Web 2.0 的出现和应用, 标签技术、网络书签、博客等成为互联网中的热门应用或服务。由于受控词表复杂而且成本高, 本体可操作性欠佳, 2004 年 8 月美国信息架构专家 Thomas Vander Wal 和 Gene Smith 提出了分众分类法 folksonomy。Folksonomy 将 Web 2.0 的特点(提倡互联网是大众的互联网, 网络的主体是人而不是机器)与分类法思想结合, 将用户信息需求和信息资源结合, 将用户对信息资源的理解与学科知识体系结合。因此, 它成为了 Web 2.0 时代的一个重要的知识组织技术。

Folksonomy 是 Thomas Vander Wal 首次提出的, 它由 folk(或 folks)和 taxonomy 两个字组成。由于 folk 字源是德文, 意思是人民、大众; 而 Folksonomy 是指用户自发性定义的平面非等级标签分类, 是由互联网用户创造的分类法。所以 Folksonomy 还被译为“自由分类法”、“大众分类法”、“通俗分类法”或“社会分类法”等。

Folksonomy 的内涵主要包括四点。第一, 信息用户参与并主导信息的组织方式; 第二, 信息用户间通过协作完成对信息的分类组织; 第三, 分类词汇无严密等级体系; 第四, 容易使用, 具有强大的知识聚合及共享功能。

Folksonomy 分为宽分众分类(broad folksonomy)和窄分众分类(narrow folksonomy)两种。它们的主要区别就是用户群的数量和异质用户的多少。顾名思义, 宽分众分类拥有大量的异质用户群, 而窄分众分类仅拥有某一专业或领域的用户群。这里的异质是指用户在知识结构和兴趣领域上具有很大的差异性, 每个用户都可能用自己的语言对社区中现有或尚未添加的内容提交独特的标签, 因此每一个内容或分类可能用相当多的标签来描述。宽公众分类法的代表是 del.icio.us, 窄公众分类法的代表是 Flickr。

Folksonomy 的分类原理是: 为向社群参与者提供一种协同构建和共享各自网络资源标签的开放式平台, 用户通过自己制定的分类标准来提交标签(tag), 由用户群体定义 tag 的频率来决定信息的组织方式。这种标签包括网络内容的分类、大意和链接地址等。tag 对每个人都是完全开放的, 用户可以自由查看并修改自己提交的标签。

作为 Folksonomy 类目的 tag, 它既不同于关键词, 也不同于目录和主题词。Tag 是在对文章概括和理解的基础上产生的, 是对内容于个人理解上的私人标注, 未必针对主题, 可能是时间、内容或与文章主题无关的词汇。因此, 标签的设定要比关键词自由和方便, 它可以从多个维度来揭示信息和知识内容。在以目录为基础的存储体系中, 目录结构必须要事前规划, 而 tag 可以不考虑目录结构并以较少的代价细化分类。Folksonomy 的类目是由用户定义一个或几个 tag 组成, 没有明确定义 tag 之间的关系, 各个 tag 之间的关系是平等的, 但是又可以根据相关性分析, 将经常一起出现的 tag 关联起来, 而产生一种相关性的分类。因此它不像传统分类法和分类表, 有多重明确的关系。Folksonomy 能解决传统分类法更新慢, 不能及时面对新出现的学科和专业术语的问题。

事实证明, Folksonomy 是一个非常有效的分类方法, 它利用“草根”的力量以一种人们喜欢的方式将网络上杂乱的信息分类组织在了一起。它的主要优势在于:

(1) 自由。任何人都可以对其感兴趣的内容进行标识, 提供一个或者多个标签; 或什么也不做, 仅浏览他人的标签。

(2) 共享。任何人对内容进行标注后, 他人可立刻看到这些标签, 如果觉得标签不合适, 还可以增加或者修改。常用的标签按字顺排列, 可以形成上面的标签 (tag cloud)。人们可以用 RSS 的方式订阅感兴趣的标签对应的资料, 或者发现志同道合者和他们的标签。

(3) 动态更新。随着人们使用不同标签标识内容信息, 标签总图中的一些标签字号会变大, 而另一些标签会逐渐“淡出”人们的视野, 这种更新使人们可随时发现当前人们关注的“热点”和信息内容的“走势”。

(4) 个性化。无论是从存储的内容, 还是分类组织的方式, 基于 Folksonomy 的数字资源系统都在反映用户的真实想法, 而不是依照提前设定好的模式, 因而它具有很强的个性化特征。

(5) 社会性。在个性化的基础上, 基于 Folksonomy 的数字资源系统给所有的用户提供了一个平台, 用户在这个平台上, 可以通过对同一内容的关注, 以及定制的服务, 分享资源。同时也能通过兴趣爱好或者邀请、添加等形式, 形成社会性的关系网络, 并与其他用户进行交流和分享。

(6) 简洁灵活性。基于 Folksonomy 的数字资源系统在界面设计、用户使用上都相当简洁和具有灵活性, 用户不需要专门的培训, 使用方便。

但 Folksonomy 的缺点也是显而易见的, 如标引词的模糊性及标引词间的差异性, 同义词的激增, 缺乏对复数的控制, 以及对语法和拼写措词的差异缺乏指引等。

最后, 再比较一下 Taxonomy 与 Folksonomy 的异同点, 如表 6.1 所示。

表 6.1 Taxonomy 与 Folksonomy 的比较

Taxonomy	Folksonomy
它是由相关专家集体编撰的权威的知识体系, 是信息构架中一个重要部分	它是群众自发性定义的平面非等级标签分类
强调放诸天下皆准的知识体系, 不会因人、因地而有所不同	是由下而上所产生的信息架构, 其内容与个人需求及社会文化情境息息相关
以利他为基础	从利己角度出发
等级式分类法便于按学科对信息进行汇集	可按照不同的属性 (分面) 查找资料, 易于为大众使用和共享
相信专家学者和专业人士, 可以花费大量财力、物力雇专人将内容进行分类	鼓励大众为了自己的需要用自己的语言标识内容, 然后通过更多人的参与, 进行人或 tag 的匹配, 或者进行有效的搜索
研制和维护成本较高	由庞大的使用者群体集体创作, 维护成本低

6.4 知识组织的技术

由知识组织的概念可知,知识组织是对知识客体进行诸如整理、加工、揭示和控制等一系列的序化、系统化的活动。具体活动包括知识加工、知识表示、知识聚类、知识重组和知识存储等。下面简要介绍在上述知识组织活动中所采用的主要技术。

6.4.1 知识加工

与信息加工过程(如计算、合并、汇总和连接等处理)不同,知识加工过程较多着眼于对知识的解析、分类、合成、整理和映射等深层处理。

知识加工的主要产品就是知识单元。知识单元是能够表达一个完整知识概念的独立单元,它也是知识组织中的基本单元。知识单元可以用主题知识单元、关键词知识单元、知识元、知识基因、知识因子、知识项、知识本体、知识概念和概念地图等各种方式来表示。在知识单元的形成过程中主要采用特征提取、概念聚类、本体、文本挖掘、知识标引和知识库等技术。

6.4.2 知识表示

知识表示是将知识对象中的知识因子和知识关联表示出来,便于人们识别和理解。常见的知识表示技术包括关系、产生式规则、框架、语义网络和本体等。相关方法已经在第5章中介绍过。

6.4.3 知识聚类

知识聚类是指将知识按一定的聚类标准分门别类地加以类集和序化的过程。它的基本原理是“事以类聚”,即根据事务的不同属性,将属性相同或相近的事务集中在一起,将属性不同的事务区分开来。按所采用的聚类标准的不同,聚类技术可分为学科聚类、主题概念聚类、知识主体聚类、用途聚类和时空聚类等。

6.4.4 知识重组

知识重组就是知识的重新整合,通过对知识及其关联的整序与加工,使知识的内部结构发生改变,然后重新加以组合,产生新的知识产品。知识重组的目的是为用户克服因知识分散而造成的检索困难提供索引指南;为人们提供经过加工整序后的精练知识;为便于用户理解和吸收知识,提供评价性和解释性的知识。

知识重组具有多种类型:按知识组分可分为文献单元组分重组和知识单元组分重组,

按知识内容可分为知识因子重组和知识关联重组,按知识连接的方式可分为有串联的知识重组和无串联的知识重组,按使用目的可分为主题性知识重组和问题性知识重组,按内容成分可分为单学科知识重组和多学科知识重组,按服务性可分为主动性的知识重组和被动性的知识重组。

知识重组能整合文本、图形、图像、音频和视频等不同媒体的知识,也能整合政府、科技和商业等属于不同系统的知识。同时,知识重组更加注重知识间的逻辑关系。

知识重组技术包括聚类分析、关联分析和自动摘要等文本挖掘和多媒体挖掘技术,以及传统的分类技术、主题标引技术、知识标引技术和文摘索引技术等。

6.4.5 知识存储

知识因能存储而得以积累和延传。知识库和知识仓库是最常见的存储技术。

知识库包括事实、规则和概念,事实是对基本信息的描述,规则是从专家们的经验中抽出来的知识,概念分为信念和常识。

知识仓库是以多行业、多类别数据仓库组成的一个集合,它涉及众多行业和众多层次的单位,在形式上包括文字、影像和图形等以多媒体形式具体存在的表现形式,也包括以某种理论、假想算法和推论形式存在的抽象的东西,是一个比较庞大的知识综合体。

知识库一般用于存放领域知识,在规模上比较小,知识的种类也比较单一;而知识仓库是面向主题的,包含的数据知识资源种类多、数量大,适用的行业也多,具有强大的知识支持辅助决策功能。



思考与讨论题

- (1) 请简述信息组织与知识组织在内涵上的异同点。
- (2) 主题图中的主题、关联和事件在知识组织中分别起什么作用?
- (3) 请选择某一领域,利用你所熟悉的本体构建方法创建领域本体。
- (4) 请分析 Taxonomy、Folksonomy 和 Ontology 三者之间的关系。



参考文献

- [1] 冷伏海,徐跃权,冯璐. 信息组织概论[M]. 第2版. 北京:科学出版社,2008.
- [2] 蒋永福. 论知识组织[J]. 图书情报工作, 2000, 6: 5-10.
- [3] 付小红. 论知识组织的原则[J]. 情报资料工作, 2001, 5: 11-15.
- [4] 李家清. 知识组织方法及策略研究[J]. 图书情报工作, 2005, 49(5): 41-44.

- [5] WordNet[EB/OL]. <http://wordnet.princeton.edu/>.
- [6] Novak J D. Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools[M]. London: Routledge, 2009.
- [7] Empolis[EB/OL]. The Topic Maps Handbook. Empolis GmbH. <http://www.empolis.com>.
- [8] 岳爱华, 孙艳妹. Taxonomy、Folksonomy 和 Ontology 的分类理论及相互关系[J]. 图书馆杂志. 2008, 27(11): 21-24.
- [9] 王忠红. 知识组织工具的发展和趋势[J]. 图书情报知识, 2009, 132: 97-102.
- [10] 戴建陆, 金涛. 网络信息组织的分类法评析[J]. 情报探索, 2008, 1: 6-8.
- [11] 温有奎. 基于“知识元”的知识组织与检索[J]. 计算机工程与应用, 2005, 1: 55-57.
- [12] 文庭孝. 知识单元的演变及其评价研究[J]. 图书情报工作, 2007, 51(10): 72-76.

第 7 章

信息与知识的检索

内容提要

信息和知识被组织起来后,检索就是根据用户的需求将相关信息和知识找出的过程。本章首先介绍了信息检索的概念及经典的检索模型。然后,介绍了知识检索的概念、机理和评价标准,重点介绍了知识检索模型包括基于概念图的检索模型、基于主题图的检索模型和基于本体的检索模型,以及基于知识的检索模型,包括分类检索模型、多维认知检索模型和分布式检索模型。最后,介绍了检索结果的可视化方法和主要技术。

本章重点

- 经典信息检索模型
- 知识检索模型与方法
- 检索结果可视化方法



7.1 信息检索

在互联网时代，网络信息呈爆炸式增长趋势。在众多类型的网络信息中，如文本信息、图像信息、音频信息、视频信息和动画信息等，文本信息数量巨大。如何从海量的文本信息中准确而全面地获取所需知识，信息检索技术是其中的关键。下面对文本信息检索的过程和常用的模型予以详细介绍。

7.1.1 信息检索的概念

信息检索是将信息按一定的方式组织和存储起来，并根据用户的需要找出有关信息的过程，它涉及信息的表示、存储、组织和存取等多个环节。

检索的本质是用户的信息需求和信息集合的比较与选择，即匹配的过程。从用户需求出发，对一定的信息集合（系统）采用一定的技术手段，根据一定的线索和准则找出（命中）相关的信息的过程，就是检索。

文本信息检索泛指通过特定算法或模型从包含各种信息的文档集中查找所需要的信息或知识的过程，它是自然语言处理的一个重要应用子领域，其研究方法主要包括传统的基于统计的方法和近年来基于语义的方法两大类。基于统计的方法是应用某些统计的手段从被检索文档和高标注等级文档中查询与用户需求匹配程度最好的文档；而基于语义的方法则尝试对需求实现一定程度语法和语义的分析，即对用户输入的自然语言文本进行一定程度的理解并重新生成查询。

文本信息检索的记录可以是结构化、半结构化和非结构化的数据格式，也可以是它们的混合体。结构化的数据包含了各种可以命名的部分，并按照一定的结构对内容进行组织，如数据库中就包含了各种结构化的记录。由于数据库中有信息表，因此可以利用 SQL 语句等搜索工具很容易地进行检索。非结构化数据无法用固定的格式对其进行组织和定义，搜索工具也无法根据特定的语义通过 SQL 语句进行检索，如一本小说中的一段文字，一份报纸中的一则广告等都属于非结构化数据。还有一些数据，尽管它们以文档的形式组织，但它们具有相同的结构和语义，文档内容按照固定条目要求进行编排，这样的数据称为半结构化数据。由于有大体相同的检索条目，半结构化的数据检索相对于非结构化的数据检索方便了很多。

7.1.2 信息检索的过程

信息检索的过程本质上是信息资源与信息需求的匹配过程，通过特定的算法寻找信息资源与信息需求的交集的过程，其中信息需求即为用户的查询请求，而信息资源是信息检

索的基础,可以是包括文本、图像、视频和音频等数据的原始文档集。文档可以是一个完整的逻辑单元,比如一本词典或一辆小汽车等。它也可以是其中的一部分,比如一个自然段或者一个汽车零件等。通常将文档看作是内容的载体,是信息检索过程中的检索单元。但对这些文档不能直接进行检索,需要从中抽取文档表示的逻辑视图,以支持信息检索。

图 7.1 显示了信息检索的基本流程。由图可见,用户首先详细说明信息需求,然后系统通过分析用户的查询请求,将用户信息需求转化成查询表达式,并对查询表达式进行分析和扩展,接下来的步骤就是在文档库中匹配出相关的文档,并通过特定的排序算法对检索到的文档进行排序,并将最终的结果返回给用户。流程中的“相关性反馈”环节,可使用户按照其自身的信息需求,对结果文档集进行筛选,将筛选后的子集反馈给系统,系统可以再利用用户的选择,进行查询表达式的改进,比如通过文档中的词语对查询表达式进行扩展,以期获得用户需求的最佳表达。

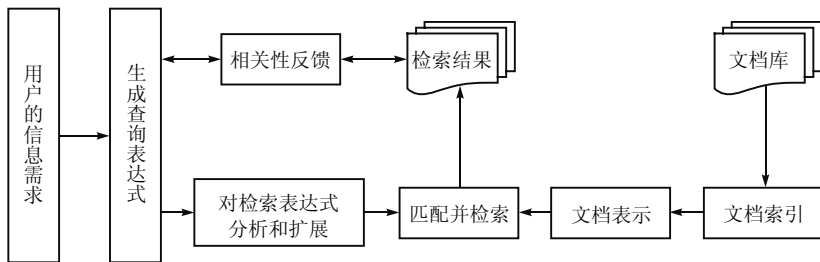


图 7.1 信息检索的基本流程

信息检索系统的主要功能包括:

(1) 用户检索的表达和扩展,根据用户的输入,生成一个查询表达式,并对查询表达式进行分析和扩展,以适合检索的要求。

(2) 文档索引,收集所要检索的半结构化或非结构化文档,整理文档并为这些文档建立索引。

(3) 文档表示,就是文档的组织形式,即如何来存储文档和表示文档的内容,建立供检索用的文档库。

(4) 匹配并检索,这是信息检索的核心部分,根据具体模型和算法从文档库中找出与用户需求相关的文档。

(5) 相关性反馈,即把检索的结果按照相关性反馈给用户,用户可以把自己对检索结果的意见反馈给系统,从而提高以后的检索质量。

7.1.3 信息检索模型

信息检索模型是描述文档和用户查询的表示形式以及它们之间的关系的框架,按照理

论基础的不同可以划分为布尔模型、向量空间模型、概率模型以及基于知识的模型等^[1]。

1. 布尔模型

布尔模型是最早提出的信息检索模型。在布尔模型中,四元组 $[D, Q, F, R(q_i, d_j)]$ 的定义为:文档 D 表示索引项的集合。用户查询 Q 表示索引项的布尔组合,用“与”、“或”、“非”连接起来,并用括弧指示优先次序。匹配 F 定义为:当且仅当一个文档满足布尔查询时,才将其检索出来,检索策略基于二值判定标准。检索算法 $R(q_i, d_j)$ 根据匹配框架 F 判定文档 d_j 是否与查询 q_i 相关,如相关,即文档 d_j 满足用户查询 q_i 的要求,便将该文档返回。

在布尔模型中,所有的索引项要么在文档中出现,要么不出现。据此,模型将所有索引项的权值全部设定为二值参数,即 $w_{i,j} \in \{0, 1\}$,当索引项 k_i 在文档 d_j 中出现时, $w_{i,j}=1$,反之 $w_{i,j}=0$ 。用户的查询 q_i 本质上是一个常规的布尔表达式。为了便于计算文档和查询的相关度,通常可以将查询的布尔表达式表示成合取向量的析取。

在布尔模型中,所有索引项的权值变量 $w_{i,j}$ 和文档 d_j 与查询 q 的相关度 $\text{sim}(d_j, q)$ 都是二值的,即 $w_{i,j} \in \{0, 1\}$ 且 $\text{sim}(d_j, q) \in \{0, 1\}$ 。查询 q 被表述成一个常规的布尔表达式,为方便计算查询 q 与文档 d 的相关度,一般将查询 q 的布尔表达式转换成析取范式 q_{DNF} ,用 q_F 来表示任意合取分量,用二值变量 $p_i(d_j)$ 表示索引项 t_i 是否在文档 d_j 中出现的值,用二值变量 $p_i(q_F)$ 表示查询合取分量 q_F 中索引项 t_i 是否出现的值。那么 $\text{sim}(d_j, q)$ 定义为

$$\text{sim}(d_j, q) = \begin{cases} 1 & \text{如果 } \exists q_F \mid (q_F \in q_{\text{DNF}}) \wedge (\forall t_i, p_i(d_j) = p_i(q_F)) \\ 0 & \text{其他} \end{cases} \quad (7.1)$$

如果 $\text{sim}(d_j, q)=1$,则文档 d_j 与查询 q 相关,否则不相关。

在布尔模型中,文档与查询是否相关是二值的,也就是说只存在相关和不相关两种情况,无法对文档和查询部分匹配的情况进行描述,这样的完全匹配方式会导致返回的文档要么量太多,要么量太少,并且无法对返回的文档按照相关度大小进行排序。即便如此,由于布尔模型简单且容易理解,经过某种训练的用户可以容易地写出布尔查询式并通过布尔查询式方便地控制查询结果,因而得到了广泛的应用,目前仍然是实际应用中最常用的信息检索模型。

2. 向量空间模型

向量空间模型(vector space model)是近年来应用较多的信息检索模型之一,它是由 Gerard Slaton 等人在 1958 年提出的。这个模型对于查询与文档的相关度有较强的可计算性和可操作性,并且已经被广泛地应用于文本检索、自动文摘、关键词自动提取、文本分类和搜索引擎等信息检索领域的各项应用中,取得了较好的效果。

在向量空间模型中,文档和查询均被看成由索引项构成的向量。例如,对于有 n 个不

同索引项的系统, 文档 \mathbf{d}_j 可以表示成 $\mathbf{d}_j=(t_1, t_2, \dots, t_n)$, 其中, 索引项 $t_k (1 \leq k \leq n)$ 常常被赋予一个权值 w_k 来表示它在文档中的重要程度, 但这个权值并非像布尔模型的索引项权值那样是二值的, 而是根据索引项对文档表示的贡献大小设定的一个大于零的值。在向量空间模型中, 一般用向量 $\mathbf{d}_j=(w_1, w_2, \dots, w_n)$ 来表示文档, 其中 $w_i (1 \leq i \leq n)$ 表示索引项 t_i 的权值。索引项的权值大小是人为赋予的, 主观性较强。为了比较客观地反映索引项对文档内容表示的贡献大小, 一个典型的确定该权值的方法是运用tf-idf公式, 即 $w_{kj}=\text{tf}_{kj}/\text{df}_k=\text{tf}_{kj} \times \text{idf}_k$ 。其中, tf_{kj} 为索引项 t_k 在文档 \mathbf{d}_j 中出现的频率, 称为索引项频率 (term frequency); df_k 则是文档集 D 中出现索引项 t_k 的文档数 (document frequency); idf_k 为 df_k 的倒数, 称为逆文档频率 (inverted document frequency)。

尽管实词一般都可以作为索引项, 但是众所周知, 一个文档中的各个实词在表达文档的含义时, 所起的作用是不尽相同的。因此, 对一个模型来说, 决定一个索引项对文档含义描述的贡献程度是一个十分重要的问题, 要从根本上解决这个问题是比较困难的。目前, 人们主要用索引项的一些容易度量和评估的属性来评价其对文献的内容描述的贡献。这主要从两个方面来考量: 第一是这个词对描述文档内容的能力大小, 这一点在向量空间模型中表现为tf值的大小; 其二是这个词区分其所在文档与其他文档的能力, 这一点在向量空间模型中表现为idf值的大小。对于一个具有多篇文档的文档集, 如果一个词 w 在每篇文档中都出现, 那么用词 w 作索引项对文档与查询之间的相关度计算的贡献就不大, 因为它不能体现各个文档的差别。但是, 当词 w 只在文档集中很少的几篇中出现, 那么用 w 作为索引项就非常合适, 因为利用它能很好地将包含词 w 的几篇文档与文档集中的其他文档区分开来。

除此之外, 还需要考虑文档长度的影响。通常, 长文档相比短文档来说更易被检出, 这将导致短文档被漏检, 因而通常还要进行标准化处理。

综上所述, 得到一种典型的经过标准化处理后的索引项权值的计算公式为

$$w_{kj}=\text{tf}_{kj}/\max_j\{\text{tf}_{kj}\} \times \lg(N \times \text{idf}_k) \quad (7.2)$$

式中, N 表示文档集中的文档个数, $\max_j\{\text{tf}_{kj}\}$ 是文档 \mathbf{d}_j 中出现频率最高的索引项的频率。

在经典模型中, 假设索引项是独立的, 或者说是正交的。也就是说, 二元组 (t_i, \mathbf{d}_j) 的权值 w_{ij} 与其他二元组 (t_m, \mathbf{d}_n) 的权值 $w_{m,n}$ 是没有关系的。这个假设极大地简化了索引项权值的计算过程, 尽管这一假设有时不符合自然语言的实际情况, 但是在这个假设下, 计算权值的过程简单快捷, 因而在目前很多实用的信息检索模型中被广泛采用。事实上, 在自然语言中, 有些索引项是相互关联的, 比如当在一个文档中看到“计算机”时, 就非常有可能同时看到“科学”; 而当在一个文档中看到“土豆”时, 看到“计算机”的可能性就很小。

在给定了文档和查询的描述后,接下来的问题便是如何看待文档和查询的相关度了。在向量空间模型中,由于查询式和文档都是向量,因而此模型用文档和查询两个向量的相似度来估计文档和查询的相关性。常见的相似度计算方法有内积法和余弦向量度量法等。

1) 内积法

内积法是将查询和文档均看成向量时,即 $q=(w_{1,q}, w_{2,q}, \dots, w_{n,q})$ 和 $d_j=(w_{1,j}, w_{2,j}, \dots, w_{n,j})$,可以使用两个向量内积的大小来近似地表示两个向量之间的相关程度,如下式所示

$$\text{sim}(d_j, q) = \sum_{k=1}^n (w_{k,q} \times w_{k,j}) \quad (7.3)$$

式中, $w_{k,q}$ 是查询 q 的第 k 个索引项的权重, $w_{k,j}$ 是文档 d_j 的第 k 个索引项的权重。在内积表示法中,向量空间模型直接根据文档向量和查询向量内积的大小对文档进行排序,内积越大,文档与查询的相关度越高。

2) 余弦向量度量法

用内积法表示查询向量和文档向量的相似度时,由于内积值没有界限,因此给相似度的表示和排序带来一定的麻烦,同时会导致在计算相似度时,长文档比短文档更具有优势。而实际上文档的长短与其是否与查询相关是没有必然联系的。为了尽可能减小文档长度这个与相似度无关的因素对相似度数值的影响,人们利用向量长度对内积进行归一化,得到用向量夹角的余弦表示相似度的模型,即余弦向量度量法。事实证明,这种方法比内积表示法的效果更好。在余弦向量度量法中,文档 d_j 和查询 q 的相似度 $\text{sim}(d_j, q)$ 定义为

$$\text{sim}(d_j, q) = \cos(d_j, q) = \frac{\sum_{k=1}^n (w_{k,q} \times w_{k,j})}{\sqrt{\sum_{k=1}^n w_{k,q}^2} \times \sqrt{\sum_{k=1}^n w_{k,j}^2}} \quad (7.4)$$

式中, $w_{k,q}$ 是查询 q 的第 k 个索引项的权重, $w_{k,j}$ 是文档 d_j 的第 k 个索引项的权重。在余弦向量度量法中,向量空间模型根据文档向量与查询向量的夹角余弦的大小对文档进行排序,夹角余弦值越大,即两者之间夹角越小,就认为文档与查询相似度越高。

向量空间模型与布尔模型相比具有较大的优势,主要体现在以下方面:

- (1) 向量空间模型索引项权重的算法提高了检索的性能,改进了检索效果。
- (2) 向量空间模型采用了部分匹配的策略,使得模型能够检索出与用户查询要求接近的文档,检索的结果文档集更接近用户的检索需求。
- (3) 向量空间模型采用一定的相似度计算方法,可以根据结果文档与查询式的相似度进行排序,从而有效地控制了返回文档的数量和质量。

但向量空间模型的不足之处在于,在构建向量空间模型时,假设任意索引项之间是

相互独立的,即在考虑索引项权重 w_{ij} 时并没有考虑其他索引项对它的影响,实际上不符合现实的语言环境。从这一点上来说,向量空间模型无法揭示索引项之间的关系,因而向量空间模型在理论上还是不够完善的。尽管如此,在实际应用中采用独立性假设计算索引项权值还是取得了不错的效果,并且计算量比较小。由于充分利用索引项之间的关联性来计算权值是很困难的,所以可通过查询扩展和相关性反馈等技术,改进查询和文档的表示,来进一步改善向量空间模型的效果。

3. 概率模型

概率模型是在布尔逻辑模型的基础上为解决检索中存在的一些不确定性而引入的,它试图在概率论的框架下解决信息检索的问题。概率模型是信息检索的重要模型之一。信息检索过程中具有的不确定性是概率模型应用到信息检索中的重要前提。在概率模型的实际应用中,为了得到模型的参数,不同的模型会有不同的假设条件,通常还会与其他的方法相结合。

信息检索系统内存在很多的不确定性,如对某一信息需求既没有唯一的查询式,也没有明确的定义和判定标准,即文档与查询是否“相关”(文档是否能满足用户的需求)。基于上述原因,Maron 和 Kuhns 在 1960 年提出了第一概率检索模型;1976 年 Robertson 和 Sparck Jones 等在此基础上进行改进提出了第二概率检索模型;之后,Turtle、Fuhr 和 Roberston 又提出了统一化模型,即第三概率检索模型,提高了文献的排序精度。

第二概率模型针对给定的一个用户查询,假设存在一个理想文档集 R ,它只是包括完全相关的文档而不包括其他不相关的文档。这样,信息检索的过程可以被看成是描述理想文档集的过程,把查询处理看作是对理想结果文档集属性的处理。当然,我们并不能确切地知道这些属性到底是什么,只知道能够利用索引项来刻画这些属性。

概率模型基于以下理论:给定一个用户的查询式 q 和文档集合中的文档 d_j 的概率模型来估计用户查询 q 与文档 d_j 相关的概率。同时,概率模型基于如下假设:

(1) 文档 d_j 与一个查询式 q 的相关性和文档集合中的其他文档是没有关系的,这称为概率模型的相关性独立原则。

(2) 在文档和查询中的索引项权重都是二元的。

(3) 文档相关性是二值的,即只有相关和不相关两种。也就是说,一篇文档要么属于理想文档集,要么不属于理想文档集。

正是由于这些假设,概率模型也称为二值独立检索模型(BIR, Binary Independent Retrieval)。

下面给出概率模型的定义。将文档 d_j 和查询式 q 的相似度 $\text{sim}(d_j, q)$ 定义为文档与查询相关的概率和文档与查询不相关概率的比值。文档 d_j 和查询式 q 的相似度 $\text{sim}(d_j, q)$ 计算公

式为

$$\text{sim}(d_j, q) = \frac{P(R|d_j)}{P(\bar{R}|d_j)} \quad (7.5)$$

式中, R 表示相关文档集合, \bar{R} 表示不相关文档的集合, $P(R|d_j)$ 表示文档 d_j 与查询式 q 相关的概率, $P(\bar{R}|d_j)$ 表示文档 d_j 与查询式 q 不相关的概率。

与其他模型类似, 概率信息检索的目的是估计 $\text{sim}(d_j, q)$, 即文档 d_j 对检索式 q 来说被用户判断为相关的概率。基于概率模型的检索过程如下:

首先, 将文档看成二值向量, 即 $d_j = \{w_{1,j}, w_{2,j}, \dots, w_{n,j}\}$, 其中 n 是所有索引项的个数。如果文档中出现相应索引项 t_i , 则 $w_{i,j} = 1$, 反之 $w_{i,j} = 0$ 。然后, 针对每一篇文档 d_j 计算 $P(R|d_j)$ 和 $P(\bar{R}|d_j)$ 来决定它是否与查询 q 相关。由于无法直接估计 $P(R|d_j)$ 和 $P(\bar{R}|d_j)$ 的值, 因此要用已知的量来进行估计。相关文档集合 R 可以由初始的猜测得到, 则概率模型中 $P(R)$ 和 $P(\bar{R})$ 以及 $P(R|d_j)$ 和 $P(\bar{R}|d_j)$ 是可计算的, 因此根据贝叶斯公式 $P(a|b) = P(b|a) \cdot P(a) / P(b)$ 可以得到

$$P(R|d_j) = \frac{P(d_j|R) \cdot P(R)}{P(d_j)} \text{ 以及 } P(\bar{R}|d_j) = \frac{P(d_j|\bar{R}) \cdot P(\bar{R})}{P(d_j)}$$

将其带入式(7.5)得到

$$\text{sim}(d_j, q) = \frac{P(d_j|R) \cdot P(R)}{P(d_j|\bar{R}) \cdot P(\bar{R})} \quad (7.6)$$

式中, $P(R)$ 和 $P(\bar{R})$ 表示从整个文档集合中随机选取一篇文档是否与查询相关的先验概率。对于一个确定的文档集来说, 这两个先验概率仅与查询有关, 而与具体的文档无关(即对 $\forall d_j$, $P(R)$ 和 $P(\bar{R})$ 均相同), 由于我们只关心所有文档与查询 q 的相似度的相对大小, 所以将式(7.6)进一步简化可得

$$\text{sim}(d_j, q) = \frac{P(d_j|R)}{P(d_j|\bar{R})} \quad (7.7)$$

根据索引项之间的独立性假设, 用每个词在相关文档集合和不相关文档集合的分布情况来计算它们的相关概率, 从相关文档集合 R 或 \bar{R} 中随机选取文档 d_j 的概率 $P(d_j|R)$ 和 $P(d_j|\bar{R})$ 的定义为

$$P(d_j|R) = \prod_{i=1}^n P(t_i|R)^{g_i(d_j)} P(\bar{t}_i|R)^{(1-g_i(d_j))} \quad (7.8)$$

$$P(d_j|\bar{R}) = \prod_{i=1}^n P(t_i|\bar{R})^{g_i(d_j)} P(\bar{t}_i|\bar{R})^{(1-g_i(d_j))} \quad (7.9)$$

式中, $g_i(d_j) = w_{i,j} \times w_{i,q} \in \{0, 1\}$, n 为所有索引项的个数。式(7.8)和式(7.9)可以这样理

解：当索引项在查询和文档 d_j 中同时出现，即 $g_i(d_j)=w_{ij} \times w_{iq}=1$ 时，用索引项 t_i 在相关文档集合 R 中的某篇文档中随机出现的概率来作为其对判断查询与文档是否相关的贡献；反之，索引项不同时出现在查询和文档 d_j 中时，用索引项 t_i 不在相关集合 R 中出现的概率来作为其对判断查询和文档是否相关的贡献。将式（7.8）和式（7.9）代入式（7.7）可得

$$\text{sim}(d_j, q) = \frac{\prod_{i=1}^n P(t_i | R)^{g_i(d_j)} P(\bar{t}_i | R)^{(1-g_i(d_j))}}{\prod_{i=1}^n P(t_i | \bar{R})^{g_i(d_j)} P(\bar{t}_i | \bar{R})^{(1-g_i(d_j))}} \quad (7.10)$$

由 $P(t_i | R)$ 以及 $P(\bar{t}_i | R)$ 的定义，有 $P(t_i | R) + P(\bar{t}_i | R) = 1$ ，相应地根据 $P(t_i | \bar{R})$ 和 $P(\bar{t}_i | \bar{R})$ 的定义有 $P(t_i | \bar{R}) + P(\bar{t}_i | \bar{R}) = 1$ ，将两者带入式（7.10），并且两边取对数可以得到

$$\text{sim}(d_j, q) = \sum_{i=1}^n g_i(d_j) \lg \frac{P(t_i | R) \cdot (1 - P(t_i | \bar{R}))}{P(t_i | \bar{R}) \cdot (1 - P(t_i | R))} + \sum_{i=1}^n \lg \frac{1 - P(t_i | R)}{1 - P(t_i | \bar{R})} \quad (7.11)$$

在式（7.11）中， $\sum_{i=1}^n \lg \frac{1 - P(t_i | R)}{1 - P(t_i | \bar{R})}$ 与文档 d_j 无关，对所有文档来说， $\sum_{i=1}^n \lg \frac{1 - P(t_i | R)}{1 - P(t_i | \bar{R})}$ 的值都是一样的，相似度排序只关心不同文档与查询相似度的相对大小，这样与文档无关的常量可以省略。因此，式（7.11）可以进一步简化为

$$\text{sim}(d_j, q) = \sum_{i=1}^n g_i(d_j) \lg \frac{P(t_i | R) \cdot (1 - P(t_i | \bar{R}))}{P(t_i | \bar{R}) \cdot (1 - P(t_i | R))} \quad (7.12)$$

这就是概率模型中用于计算文档和查询相似度的公式。

概率模型的优点主要包括：

（1）有严格的数学理论基础。概率模型的文档与查询相似度值计算公式以严格的数学理论以及严格的推导为依据获得的，因此概率模型为人们提供了一种进行检索决策的数学理论基础。

（2）概率模型可以采用相关反馈原理，可开发出理论上更为可靠的方法。

（3）概率模型中没有用到对用户的查询技术要求比较高的布尔逻辑方法，同时可以将文档按照它们相关概率的递减顺序排序。

概率模型的缺点主要包括：

（1）在模型中假设索引项的权值是二值的，没有考虑不同索引项在文档和查询中的不同权重，这个假设产生的问题在布尔模型中已经作了说明。

（2）需要在开始之前就将文档分成相关文档集合和不相关文档集合，这往往给用户带来很大困难或者区分准确度很低。

4. 扩展布尔模型

扩展布尔模型通过建立数学模型来表示布尔逻辑和通过加权机制改进索引项的权值两个方面对布尔模型进行改进。

1) 基本模糊集合模型

布尔模型的理论基础是布尔逻辑和经典集合论。扩展布尔模型的改进之一，就是采用模糊集合论作为理论基础，用函数计算代替布尔逻辑计算。

令索引项 t_i 在文档 d_j 中的权值为 $w_{i,j}$ ($0 \leq w_{i,j} \leq 1$)，在查询中出现的索引项 t_i 的权值为 $w_{i,q}$ 。在这个模型中，可以认为 $w_{i,q} = w_{i,j}$ 。当索引项权值退化为 0 或 1 时，模糊检索模型与传统检索的布尔模型是完全兼容的。对于一个标准布尔系统，可以用赋值表来处理表达式；对于一个包含多个运算符的检索式，可先处理深层的检索子式，再逐层递归处理。布尔检索的检索式与文档 d_j 相似度计算公式如表 7.1 所示。

表 7.1 布尔检索的检索式与文档 d_j 相似度计算公式

布尔检索式	赋值公式
$\text{sim}(d_j, t_m \text{ and } t_n)$	$\min\{w_{m,j}, w_{n,j}\}$ $w_{m,j} \cdot w_{n,j}$
$\text{sim}(d_j, t_m \text{ or } t_n)$	$\max\{w_{m,j}, w_{n,j}\}$ $w_{m,j} + w_{n,j} - w_{m,j} \cdot w_{n,j}$
$\text{sim}(d_j, t_m \text{ and not } t_n)$	$w_{m,j} \cdot (1 - w_{n,j})$

2) 扩展模糊集合模型

在基本模糊集合模型中，没有对查询式中的检索词赋予权值，而是假设查询中的索引项权值和文档中的索引项权值相等。扩展模糊集合模型针对这一点进行了改进，为文档中的词和检索式中的词赋予了不同的权值。

在向量空间模型中，查询和文档都采用同一种结构表示，所以没有必要区分查询索引项和文档索引项。但是在布尔模型和模糊模型中，查询的结构化要求区分查询索引项和文档索引项的赋值过程。在布尔模型中，给查询索引项赋权值，必须保留布尔模型的特性：

- (1) 兼容性，指与查询索引项赋值为 0 或 1 时的结果兼容。
- (2) 一致性，指查询式 $(t_1 \text{ or } t_2) \text{ and } t_3$ 与查询式 $(t_1 \text{ or } t_3) \text{ and } (t_2 \text{ and } t_3)$ 检索出相同的文档。
- (3) 独立性，指对查询式进行分段处理，不会影响整个查询式的检索结构。

一种有效的方式是将查询索引项的权值与文档索引项的权值相乘，即 $\text{sim}(d, t_i) = w_{i,d} \times w_{i,q}$ ，其中 $w_{i,d}$ 和 $w_{i,q}$ 分别表示索引项 t_i 在文档和查询中的权值， $\text{sim}(d, t_i)$ 表示当查询只有索引项 t_i 时，查询和文档 d 的相似度。

5. 统计语言模型

统计语言模型 (SLM, Statistical Language Mode) 试图通过统计学和概率论理论对自然语言进行建模, 从而获取自然语言中的规律和特性, 以解决语言信息处理中的特定问题。

语言 (或者说文档) 就是字母表上的某种概率分布, 该分布反映了任何一个字母序列成为该语言的一个句子 (或其他任何语言单元) 的可能性, 这个概率分布称为语言模型 (language model)。对于任何一个句子 $S = s_1, s_2, \dots, s_n$ (其中 n 为句子的长度), 将生成句子的过程看成一个马尔可夫过程, 则其在文档集中出现的概率为

$$P(S) = P(s_1, s_2, \dots, s_i) = \prod_{i=1}^n P(s_i | s_1, s_2, \dots, s_{i-1}) \quad (7.13)$$

有了这个定义之后, 如何根据给定的数据集 (训练语料库) 来估计概率 $P(s_i | s_1, s_2, \dots, s_{i-1})$ 就成为统计语言模型中的关键问题。由于不可能有足够的数据库来估计 $P(s_i | s_1, s_2, \dots, s_{i-1})$, 因此根据马尔可夫假设, 可得到 n 元模型 (n -gram 模型), 即认为一个词的出现与否仅仅与其前面的 $n-1$ 个词有关, 即

$$P(s_i | s_1, s_2, \dots, s_{i-1}) \approx P(s_i | s_{i-1}, s_{i-2}, \dots, s_{i-n+1}) \quad (7.14)$$

当 $n=1$ 时, $P(s_i | s_1, s_2, \dots, s_{i-1}) \approx P(s_i | s_i) \approx P(s_i)$, 这时的模型称为一元语言模型 (unigram)。它假设各个词之间是相互独立的, 即一个词出现的概率与其他词无关。

当 $n=2$ 时, $P(s_i | s_1, s_2, \dots, s_{i-1}) \approx P(s_i | s_{i-1})$, 这时的模型称为二元语言模型 (bigram)。它假设索引项出现的概率仅与其前一个词出现的概率有关。

当 $n=3$ 时, $P(s_i | s_1, s_2, \dots, s_{i-1}) \approx P(s_i | s_{i-1}, s_{i-2})$, 这时的模型称为三元语言模型 (trigram)。它假设索引项出现的概率仅与其前两个词出现的概率有关。

基于统计语言模型的检索模型的基本思想是对于每一篇观察到的文档, 假设其对应着一个语言模型, 并根据这篇观察到的文档估计这个语言模型。同时假设检索的过程就是由文档的语言模型产生查询的过程。在这个过程中, 计算语言模型生成查询的概率成为基于语言模型的检索模型的核心, 即检索最感兴趣的文档 $P(D|Q)$, 在 $P(D|Q)$ 上使用贝叶斯法则, 并略去对文档排序不造成影响的因素, 得到

$$P(D|Q) \propto P(Q|D)P(D) \quad (7.15)$$

式中, $P(D)$ 是一篇文档符合查询的先验概率, $P(Q|D)$ 刻画了在给定文档 D 时, 查询 Q 生成的可能性。实际上这个概率也就揭示了文档 D 符合给定的查询 Q 的程度 (即文档 D 与给定的查询 Q 的相似度)。

$P(D)$ 显示了利用先验概率来改善查询的效果的可行性, 比如可以考虑从文档的内部

结构、对文档在整个网络中所处的链接环境分析等获得先验概率的知识。假设这个先验概率对文档集合中的所有文档都是相同的,于是得到的基于语言模型的检索模型为

$$P(D|Q) = \prod_{w \in Q} P(w|D) \quad (7.16)$$

从表面上看,基于统计语言模型的检索模型与传统的向量空间模型非常不同。它比较简单,仅仅利用了词的频率信息,然而在实验中确有很好的效果。

基于统计语言模型的信息检索模型的优/缺点包括:

与传统的向量空间模型相比,基于统计语言模型的信息检索模型取得了较好的检索效果。基于统计语言模型的信息检索模型由于具备数学理论基础坚实,概念模型简洁并且在实际测评中相对于传统的向量空间模型也确实获得了更好的检索效果等特点,近年逐渐引起研究人员的注意。目前,基于统计语言模型的信息检索模型已经成为信息检索研究的一个新方向。改进的基于统计语言模型的信息检索模型也不断涌现,如触发语言模型和基于主题语言模型等。但其主流的思想都是将自然语言处理中的技术(如词义消歧、句法分析和指代消解等)引入到统计语言模型中,以提高检索效果。

同时,统计语言模型也存在一些需要进一步研究与解决的问题。在统计语言模型中,由于将查询看作是由文档模型抽样生成的样本,因此,对于传统的信息检索模型中使用的查询扩展技术,无法在该模型中找到充分的理论根据。

目前用于信息检索的统计语言模型基本上还只限于 n -gram 的线性词汇统计模型,这只是单词级别上的检索。因此,利用加入了句法的结构化统计语言模型进行信息检索将是一个可行的研究方向。

6. 潜在语义索引模型

自然语言文本中的词汇(术语)具有一词多义(polysemy)和一义多词(synonymy)的特点。只用索引项来表示查询和文档时,由于一词多义,基于精确匹配的检索算法会返回许多与用户查询不相关的文档;由于一义多词,基于精确匹配的检索算法又会遗漏许多与用户查询相关的文档。

为解决上述问题,以 Belcore 和 Dumais 为首的研究小组提出了一种称为潜在语义索引(LSI, Latent Semantic Indexing)的模型,该模型试图绕过自然语言理解,用统计的办法达到既反映术语间内在的相关性又改善检索效果的目的。

LSI 被证明是一种比在 Salton 的 SMART 系统中使用的传统向量空间技术性能更好的信息检索向量空间技术。

潜在语义索引模型原理如下所述。

首先,以索引项为行、文档为列形成一个大矩阵,设矩阵名为 \mathbf{X} 。

假设文档集合中索引项个数为 t , 文档个数为 d , 则矩阵 \mathbf{X} 共有 t 行 d 列, 矩阵的元素 x_{ij}

为索引项 t_i 在文档 d_j 中出现的频度,有时 x_{ij} 还加入了索引项的权重。索引项的权重说明了所有索引项在语义空间中的重要性是不相同的。定义索引项权重的原则是,识别文档与语义能力强的索引项的权重高于识别能力弱的索引项,目前定义索引项权重的方法包括传统的tf-idf方法、term熵(entropy)方法、微软提出的Okapi方法等。由于任意一个文档总是由有限个词汇组成,不大可能会有文档集中出现所有 t 个索引项,所以 X 必是一个稀疏矩阵。根据奇异值分解定理,矩阵 X 可以分解为三个矩阵 T_0 、 S_0 、 D_0^T (D_0 的转置)的积,即

$$X = T_0 S_0 D_0^T \quad (7.17)$$

式中, T_0 为 $t \times r$ 矩阵, S_0 为 $r \times r$ 对角矩阵, D_0^T 为 $r \times d$ 矩阵。其中 T_0 和 D_0 的列向量都是正交归一化的, r 是矩阵 X 的秩。

取正整数 k , $0 < k < r$,在 S_0 中,仅考虑其中值最大的 k 个奇异值(即对表达文档含义贡献最大的前 k 个词),取 S_0 中相应的 k 阶对角矩阵、 T_0 中相应的 k 列、 D_0^T 中相应的 k 行,最终得到新的矩阵 T ($t \times k$ 阵)、 S ($k \times k$ 对角矩阵)和 D^T ($k \times d$ 阵)。然后,进行奇异值分解反运算,得到新矩阵 \hat{X} ,即

$$\hat{X} = T S D^T \quad (7.18)$$

式中, \hat{X} 即为经过优化的语义结构矩阵, T 为索引项矩阵, D 为文档矩阵。奇妙的是, \hat{X} 在最小二乘意义下是 X 的最佳近似。这样,实际上得到了一个“降维”的途径。根据矩阵 T 和 D ,可以得到索引项及文档在 k 维语义空间内的坐标向量。下面要说明 T 、 S 、 D 三个矩阵在文档检索中的重要应用价值。

给定矩阵 X ,基于 X 可以提出以下三类与文档检索密切相关的问题:

- (1) 索引项 t_i 与 t_j 有多相似?即索引项的类比和聚类问题。
- (2) 文档 d_i 与 d_j 有多相似?即文档的类比和聚类问题。
- (3) 索引项 t_i 与文档 d_j 有多相似?即文档与索引项的匹配问题。

得到 \hat{X} 后,可以用 \hat{X} 来进行这三类比较。下面说明具体的比较过程。

1) 比较两个索引项

首先做“正向”乘法,即

$$\hat{X} \times \hat{X}^T = T \times S \times D^T \times D \times S \times T^T = T \times S^2 \times T^T \quad (7.19)$$

因为 D 是正交归一的,所以 $D^T \times D = I$,所得新矩阵的第 i 行第 j 列元素值的大小表明了索引项 t_i 与 t_j 的相似程度。

2) 比较两个文件

做“逆向”乘法,即

$$\hat{\mathbf{X}}^T \times \hat{\mathbf{X}} = \mathbf{D} \times \mathbf{S} \times \mathbf{T}^T \times \mathbf{T} \times \mathbf{S} \times \mathbf{D}^T = \mathbf{D} \times \mathbf{S}^2 \times \mathbf{D}^T \quad (7.20)$$

因为 \mathbf{T} 也是正交归一的, 所以 $\mathbf{T}^T \times \mathbf{T} = \mathbf{I}$, 所得新矩阵的第 i 行第 j 列元素值的大小表明了文档 \mathbf{d}_i 与 \mathbf{d}_j 的相似程度。

3) 比较一个文件与一个索引项

所有文档与索引项的相似度大小矩阵恰巧就是 $\hat{\mathbf{X}}$ 本身, 它的第 i 行第 j 列元素值大小表明了索引项 t_i 与文档 \mathbf{d}_j 的相似程度。

利用潜在语义索引模型进行信息检索的过程如下:

首先用户为了获得自己需要的文档, 提交查询 \mathbf{q} 。传统的基于关键词的用户查询必须是分离的索引项序列, 而潜在语义索引模型允许用户提交自然语言形式的查询。在处理查询时, 与对文档的处理一样, 潜在语义索引根据用户查询式中各词汇的出现频率生成查询式向量 \mathbf{q} , \mathbf{q} 中第 i 个元素的数值表示第 i 个索引项在查询 \mathbf{q} 中出现的次数。由于前面已经对索引项—文档矩阵 \mathbf{X} 进行了截断奇异值分解, 因此对于 \mathbf{q} 也要进行类似处理, 使其可以与文档矩阵 \mathbf{D} 进行比较。用下式对 \mathbf{q} 进行处理, 即

$$\mathbf{D}_q = \mathbf{q}^T \mathbf{T} \mathbf{S}^{-1} \quad (7.21)$$

式中, \mathbf{D}_q 即为查询式 \mathbf{q} 在 k 维语义空间内的坐标向量。词汇、文本、查询式三者的坐标向量构成了所需的潜在语义空间。得到了查询式的坐标向量 \mathbf{D}_q 后, 与文档矩阵 \mathbf{D} 的每一个行向量 \mathbf{d}_j 进行比较, 就可分别计算出每篇文档 \mathbf{d}_j 与查询式的相关程度, 最简单常用的方法是计算两者之间的夹角余弦值, 余弦值越大(即两个向量的夹角越小), 相似度就越大。将该相似度计算方法运用到LSI中, 根据下式来计算每篇文档与查询式的相似度

$$\text{sim}(\mathbf{d}_j, \mathbf{q}) = C(\mathbf{d}_j, \mathbf{q}) = \frac{\sum_{i=1}^k q_i \cdot t_{i,j}}{\sqrt{\sum_{i=1}^k (q_i)^2} \sqrt{\sum_{i=1}^k (t_{i,j})^2}} \quad (7.22)$$

式中, $C(\mathbf{d}_j, \mathbf{q})$ 为查询向量 \mathbf{q} 在 k 维空间内的坐标向量 $(q_1, q_2, \dots, q_k)^T$ 与某文档向量 \mathbf{d}_j 在 k 维空间内的坐标向量 $(d_{1,j}, d_{2,j}, \dots, d_{k,j})^T$ 之间的夹角余弦。

对所有文档都计算与查询的相似度值(即文档向量与查询向量的夹角余弦值)并进行排序之后, 设定一个阈值 S_{\min} , 将所有相似度大于等于 S_{\min} 的文档返回, 便可产生一个满足用户要求的检索结果文档集。通过设置不同的阈值, 可以调整返回文档的数量和文档与用户查询相似度的大小。

潜在语义索引模型的特点如下:

(1) 在潜在语义索引模型中, 实际上是将高维语义空间中的文档向量(索引项向量)

投影到低维语义空间的潜在语义空间中,达到将原来在高维语义空间中比较稀疏的向量压缩到潜在语义空间中不再稀疏的向量。即使两个文档原来没有共同索引项(或索引项完全不相同),在潜在语义索引模型中仍然可能找到它们之间比较有意义的关联值。

(2) 潜在语义索引模型是一种全新的半智能型模型,相对于传统的索引项匹配法,它是对旧有信息检索模型的一种改革。目前,它仍处于发展阶段,国内外的许多学者都在研究、开发和改善它。潜在语义索引模型在专业领域的检索方法上的效果尤其突出,一些信息检索系统也已采用潜在语义索引模型进行工作,如互联网化学资源导航系统(ChIN, The Chemical Information Network, <http://www.chinweb.com>)等。

多数情况下,潜在语义索引模型的性能要好于向量空间模型,它适用于词汇异度较高的文档集合。然而,从应用的角度来说,潜在语义索引模型计算量太大,并且目前潜在语义索引模型的理论基础还不是很完善。例如,潜在语义空间维数 k 的取值主要依靠人的经验和实际检验来确定,文档长度如何选取直接影响潜在语义提取以及再现语义的效果,目前也只能依赖经验而定。此外,由于中文信息的特殊性,潜在语义检索模型在中文信息检索中的效果不如在英文中的好,把潜在语义检索模型成功地应用于中文信息检索,还需要大量的研究。

7.2 知识检索

知识检索是在现有信息检索技术上发展而来的,但与信息检索不同的是,知识检索强调的是基于知识的、语义上的匹配。下面分别对知识检索的概念、特征、机理和效果评价进行介绍。

7.2.1 知识检索的概念

知识检索是指在知识组织的基础上,根据用户的需求,在知识库中检索出知识的过程,是一种基于知识组织体系,能够实现知识关联和概念语义检索的智能化的检索方式。

知识检索的基本思想就是模拟扩展人类关于知识处理与利用的智能行为和认识思维方法。知识检索综合运用信息管理科学、人工智能、认知科学及语言学等多学科的先进理论和技术,基于知识和知识组织,融合知识处理和多媒体信息处理等多种方法和技术,充分表达和优化用户需求,能高效存取文本、图像、视频和声音等媒体类型的知识源,并能准确精选用户需要的结果,因此能够满足用户在语义和知识上的需求^[2]。

知识检索与信息检索不同。知识检索强调语义,它从文档中的语义、概念出发,能够揭示文档内容的含义,而不像信息检索只是基于字面的机械匹配。知识检索提高了检索效率,减轻了用户的负担。

为帮助读者进一步理解知识检索的概念，我们从检索语言、检索模型、组织方式、存储方式等方面，对知识检索和信息检索做了比较，如表 7.2 所示。

表 7.2 知识检索与信息检索的比较

比 较 项 目	信 息 检 索	知 识 检 索
检索语言	自然语言	知识结构+自然语言
检索模型	布尔模型、向量空间模型	语义模型（如概念图、主题图、本体和语义网络等）
组织方式	索引	知识因子之间的关联和知识结构
存储方式	文档库	知识库
匹配/推理	部分匹配，最佳匹配/归纳推理	部分匹配，最佳匹配/检索+演绎推理/归纳推理，关联推理，类比推理
检索方式	关键词、目录	机器学习、人工智能
检索结果	文本段落或文档	知识因子/知识单元
检索效率指标	查全率、查准率、检索时间	内容相关性

7.2.2 知识检索的特征

知识检索的基本特征是对包含在信息记录中的知识和知识关联进行分析，运用知识处理技术和知识组织技术，实现基于语义理解的智能化查询。

知识检索具有以下一些特征：

（1）知识检索基于某种具有语义模型的知识组织体系。知识组织体系是实现知识检索的前提和基础，而知识检索则是基于知识组织体系的结果。

（2）拥有知识库。在知识检索中，文档内容和查询内容都以概念和范畴等知识形式存储在知识库中，用来匹配用户的查询式，推理出满足用户需求的新信息。

（3）支持语词、语义内容的处理，可以实现同义词扩展检索和关联检索。

（4）具有概念推理和学习功能。

（5）具有记忆能力。知识检索通过记忆机制，将分析出的概念信息和范畴信息存储到知识库中，并能自动补充和更新，还能进行必要的逻辑推理。

（6）人-机接口。知识检索能根据文档内容和用户查询式构造检索要点并输入系统，其输出的是按用户要求进行加工的结果，并可以自然语言的形式提供给用户。

7.2.3 知识检索机理

知识检索与信息检索的机制不同，前者更加强调语义的相关性。知识检索的实质是通过文档原文信息进行语义上的自然语言处理，找出各种概念信息，形成知识库，从概念层次上处理用户的检索和查询。利用知识检索不仅能检索包含有查询关键词的结果，

还能检索出那些与该词同属一类概念的词汇的结果。

知识检索机理如图 7.2 所示。它的检索过程如下：

在领域专家、语言学专家等的帮助下，建立知识库；采集信息源，并借助知识库中的概念网络以及各种语义词典和规则对信息源进行语义分析，把抽取的概念语义及索引信息按规定的格式存储在索引库中；获取用户查询请求，查询转换器借助知识库以及通过与用户的交互并分析查询请求，确定用户的真实查询请求，并对其进行语义分析；在概念网络和各种语义词典的帮助下，确定查询所涉及的术语及其关联，扩展和精准用户查询请求，并把查询请求转换成规定的格式；将检索请求语义与抽取的信息源语义进行匹配，并把匹配的结果经过定制处理返回给用户。

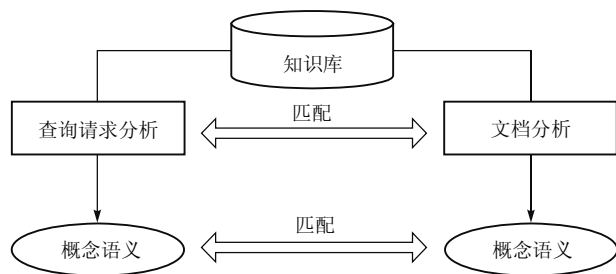


图 7.2 知识检索机理

7.2.4 检索效果评价

检索效果评价通常使用准确率（查准率）、召回率（查全率）和检索速度三个参数进行度量。准确率测量检索的准确性，它被定义为检索到的相关文档数与检索到的文档数的比值，见式（7.23）；召回率测量从数据库中检索相关信息项的能力，它被定义为检索到的相关文档数与数据库中相关的文档数的比值，见式（7.24）。速度含义是显然的，速度越高，性能越好。对海量的网络信息而言，传统的信息检索方法不能给出准确的回答。因此，准确率和召回率都不高。

$$\text{准确率 (precision)} = \frac{\text{检索到的相关文档数}}{\text{检索到的文档数}} \quad (7.23)$$

$$\text{召回率 (recall)} = \frac{\text{检索到的相关文档数}}{\text{实际相关的文档数}} \quad (7.24)$$

准确率用于回答下面的问题：“检索到的所有文档都是用户感兴趣的吗？”召回率则回答这样的问题：“所有相关的文档都检索出来了吗？”如果一个文档应该被查询检索出来，则称该文档是相关的。图 7.3 显示了文本信息检索系统查询可能得到的四种查询结果，在如图所示的四个象限中，有两个代表了所希望的结果：相关并检索到了和不相关且没有检索到。另外两个象限代表了出错情况：相关的文档应该被检索出来，但没有被

检索出来；不相关的文档应该不被检索出来，但却检索出来了。

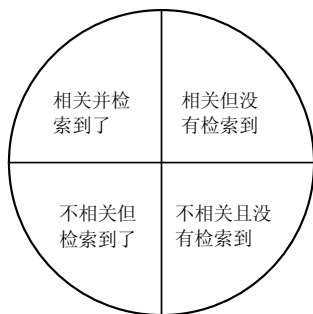


图 7.3 文本信息检索系统查询结果的度量

用户总是希望获得具有最佳准确率和召回率的结果文档集合，他们希望只检索到所有相关的文档。但是，高的准确率和召回率是很难实现的。所以在实践中，有的用户只需要一个很小的结果集，因而他更偏重于准确率；而有的用户需要一个大的结果集，那么他就会偏重于召回率，即使这个大集合中有部分结果集与查询的相似度很小。因此必须综合考虑准确率和召回率这两个指标才能准确地衡量出检索系统的性能。

知识检索强调的是基于知识的、语义层面的匹配，是在自然语言理解、计算语言学发展的基础上产生的，因此在准确率和召回率上能更好地满足用户的检索要求。

7.3 知识检索模型

知识检索模型是指集成各类知识对象和信息对象，如领域知识、用户知识、专家经验知识以及多媒体信息元素等；融合各种智能与非智能理论、方法和技术，如推理、机器学习、数据挖掘、智能搜索、智能演算以及多媒体信息处理等；以实现知识检索。

如前所述，知识检索是以知识组织体系为基础的。本书第 6 章中介绍了各种知识组织方法，下面将根据概念图、主题图和本体三种知识组织体系分别介绍各自的检索模型。此外，基于知识的检索模型，如分类检索模型、多维认知检索模型和分布式检索模型，也在本节做简要介绍。

7.3.1 基于概念图的检索模型

概念图使得传统的基于关键词的检索上升到语义检索的高度。其基本思想是：先建立相关领域的概念图，根据概念图将收集的相关信息资源进行标引和整合，形成完整的知识模型，以此作为检索的基础。用户的查询请求按照概念图转换成规定的格式，在概念图的帮助下匹配出符合条件的概念图集合返回给用户。概念图不但为知识检索奠定了

良好的知识组织基础,同时也提供了很好的检索机制。而且,概念图 also 具有良好的网络特性,能较好地解决网络信息资源的检索问题。

基于概念图的知识检索过程如图 7.4 所示,其中概念图的语义匹配是关键。概念图的匹配,既可以采取准确匹配方式,也可以采取计算两个概念图间的相似度的方式。计算相似度时,需注意概念图的相似度是不对称的,通常考量的是资源图匹配查询图的程度,而不是反之。

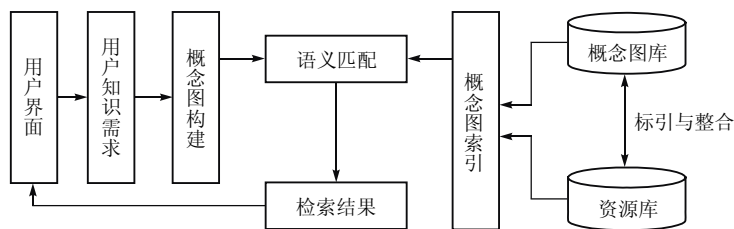


图 7.4 基于概念图的知识检索过程

基于概念图的知识检索机制主要涉及以下六个方面。

1) 标引

为了加速并简化对概念图 and 资源的搜索,可为所有在客户机和服务器上的概念图 and 资源作标引。标引内容主要包括:每个概念图中所有的概念、连接短语、特性(所有者、研讨主题等)、记录(是否隐藏)等;标引所有资源的名称和说明(图像、视频和声音等),标引文本或网页链接的内容。

2) 概念图服务器

检索需要使用概念图索引 and 服务器。服务器包括概念图服务器、索引服务器、网页检索服务器 and 语词网服务器。概念图服务器定期向一个或多个位置目录传递服务器和服务的位置信息,可以编制索引,并将索引复制到索引服务器中去。

3) 资源索引及索引服务器

被检索的概念图及资源会被编入索引,其中的资源分为概念图、图像、网页、文件和视频等类型。资源的索引信息包括作者、描述/核心问题、关键词和资源内容等。

4) 网页搜索服务器

概念图网页是概念图的一部分,可以通过互联网进行概念图网页检索。用户将网页查询请求传递给网页搜索服务器后,网页搜索服务器通过搜索互联网产生候选概念图网页,然后对其进行过滤和排序,最后产生相关网页。

5) 位置目录

位置目录是一种特殊类型的服务器,它操纵着来自概念图服务器的位置信息。当一台概念图服务器开始运行时,就在位置目录中登记一次位置信息,表明它正在运行。一

台概念图服务器可以登记一个或多个位置目录,并定期向这些位置目录传递位置信息。用户可以在互联网的任何地方对概念图服务器所发布的位置进行定位。

6) 概念图的语义和拓扑特征

概念图的命题、半等级结构、上下文关联环境等独特特性为检索奠定了很好的基础。概念图的命题特性和它们以等级排列的拓扑结构提供了足够的上下文信息以用来识别和排序与概念相关的资源。它利用其语义和拓扑特征能够比仅仅依靠一系列加权词提供更多的信息。概念图是一系列概念的二维表示形式,它使概念之间的关系更加明晰,语义更丰富。其语义拓扑特征可以帮助人们提高检索效率,最后仅把那些高度相关的内容返回给用户。

基于概念图的知识检索模型的优势主要体现在以下八个方面。

1) 上升为语义层面

由于概念图揭示了知识间及知识与资源间的语义关联,因而可以支持语义层面的检索。基于语义的检索要比基于关键词的检索更智能,检索也更为精确。

2) 检索速度快

概念图被索引的存储在索引服务器中,用户可以通过索引服务器检索相关的概念图。这种检索方式赋予概念图较强的检索能力,使得概念图知识模型用户能够在知识模型中快速准确地找到所需知识的位置。因此速度要优于传统的信息检索系统。

3) 检索的智能化

概念图中的知识组织是以概念图为知识表示形式同时还存储相关的信息资源,它是对概念及其关联的确切表述和分类,实现了知识结构与相关资源的整合。概念图采用图形匹配的技术完成检索,并以可视化的方式帮助用户进行知识导航,使用户很容易地定位并且快速地检索到他们所需要的知识点,同时得到相关资源。因此,基于概念图的知识检索和知识导航比传统的信息检索和导航更加智能化。

4) 知识形式的检索结果

基于概念图的知识检索返回的检索结果不但包含了被检索词的文档,而且检索结果仍然是以概念图形式返回的,因此可以揭示所需的概念间以及概念与资源间的关联关系。基于概念图的检索将检索结果以知识的形式返回给用户,能极大地满足用户的应用需求。

5) 检索结果中包含资源

用户可以直接利用概念图进行知识导航,或者通过输入用户提问对概念图进行检索。传统检索仅仅返回结果列表,但基于概念图的检索返回的结果中直接包含了所需资源,因而方便用户的使用。

6) 检索结果排序良好

概念图提供了强大的上下文信息,为检索结果排序奠定了较好的基础。基于概念图

的检索机制可以利用概念图中概念及其关联的权重,鉴别检索结果与主题相关程度,进一步在检索结果上提高排序效果。概念图在检索结果排序方面效果很好。

7) 检索结果易于吸收

基于概念图的检索为用户提供了包含完整的知识结构及丰富的媒体资源的可视化检索结果。这种结果更全面、更综合、更合理和更直观,有利于用户的吸收和更好地利用。

8) 准确率高

由于概念图是通过概念之间的关系来表达概念语义的能力,基于概念图的知识检索对知识、资源及用户需求进行语义分析,在检索过程中,采用的不是通过关键词进行全文检索,而是基于概念的检索。因此,检索结果的准确性好,检索的查准率高。

7.3.2 基于主题图的检索模型

主题图改变了传统意义上的匹配检索模式。它揭示了信息资源的主题概念,并将主题概念作了链接,展示了整个资源库的知识结构,从而使用户可以在浏览关于该资源库的知识结构图的前提下,进行信息资源的检索。

在主题图中,概念与概念之间通过主题进行关联,构成一定程度上的语义网络。用户首先可以从他需要或者感兴趣的主题出发来浏览整个主题图;同时,与主题相关的主题亦会被列出,用户可以根据需要选择相关主题继续浏览,以获得更多的信息,扩大有效检索范围的覆盖面。这种基于图形的可视化的检索方式不仅具有直观的特点,而且可以从可视化的知识中发现隐含知识,其交互式检索模式可以更准确地表达用户的信息需求。

基于主题图的检索方法以其直观、纵揽全局的特点,为用户提供了以下功能:

(1) 为用户提供相关知识提示,对于初次涉及某一领域的用户更可提供知识引导和提示。

(2) 帮助用户进行整体文档的结构定位,即定位浏览者目前的文档节点位置和已经浏览过的节点路径。

(3) 根据用户的浏览行为给予浏览信息导航,即在用户需要帮助时,提供可能有价值的相关概念信息。

基于主题图的检索中,条件查询是最基本的检索方式,即用户输入检索查询条件,系统在检索完成后,自动将满足条件的检索结果返回给用户。

条件查询的本质是查询项与标引项的匹配问题,即概念间相似度的度量问题。关于概念间语义的相似度计算,可以采用基于编辑距离的语义相似度算法、基于结构的语义相似度算法等,这里不做详述。基于主题图的知识检索过程如图 7.5 所示。当用户输入查询词后,通过计算查询词与主题词间的语义相似度,可以找到与查询词相关的主题,

然后将其按照与查询词的语义相关程度进行排序，这样，用户可在排序后的主题上进行浏览，直到获得需要的信息资源。

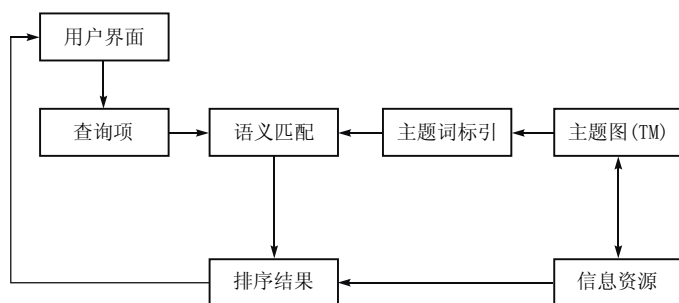


图 7.5 基于主题图的检索过程

基于主题图的条件查询，实际上相当于为资源建立了独立于实际资源的、灵活的动态索引。当用户输入任意一个关键词后，不是在整个资源查找相关信息，而是在主题域中查找相关信息，这样，可提高检索效率。

7.3.3 基于本体的检索模型

基于本体的知识检索可以描述为：在领域专家的帮助下建立领域本体，然后把收集来的数据按规定格式存储在知识库中；查询转换器按照建好的本体把查询请求转换成规定的格式，从知识库中匹配出符合条件的数据集合，其检索的结果经过定制处理后返回给用户。

基于本体的知识检索模型在资源对象的组织、描述、表示、检索和模型约束等方面都具有自己的特征，主要表现为：

(1) 在检索对象的组织上，知识检索模型利用领域本体作为组织资源的基础。首先构建一个涵盖相关领域概念及概念间关联的领域本体库作为资源描述和知识表示的工具和模型，在此基础上确定领域知识本体的主要概念和概念间的各种关系，构筑领域本体的概念模型。

(2) 在检索对象的描述上，知识检索模型借助语义标引工具，按照领域本体的概念及关联，对资源对象进行概念分析、分类、标引、描述和处理，形成机器可以理解的带有语义信息的元数据。

(3) 在模型约束上，知识检索模型的约束比较少，但要求概念和关系构成一个有向图，其关系必须是有方向的；另外，要求每一个概念/实例都需要有一个唯一的标识。

在具体应用中，不同的基于本体的知识检索系统，其工作原理不尽相同。但一般都包括人-机交互部分、知识源部分、检索匹配和本体库这四部分，如图 7.6 所示。

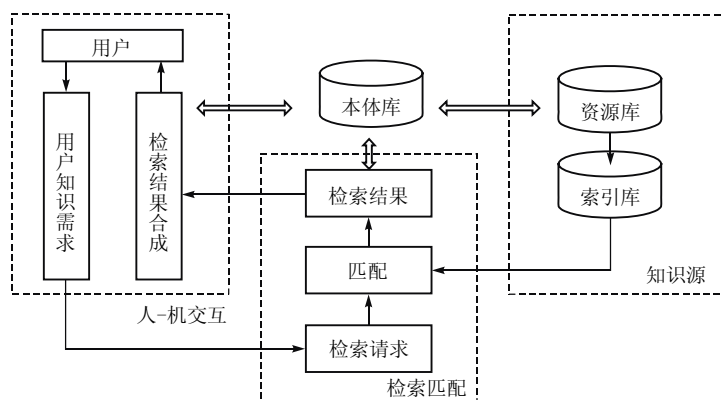


图 7.6 基于本体的知识检索模型

(1) 人-机交互部分主要是分析用户提交的知识检索请求，返回整理检索结果。其检索请求的分析主要是依据本体知识，分析用户的真实检索意图，形成规范、准确的检索请求，然后提交给检索匹配模块。

检索请求的分析主要分为以下步骤：

- ① 对检索请求进行预处理，提取需要检索的关键字（词）；
- ② 借助本体并在必要时通过与用户再次交互，判断检索请求中关键字（词）的领域、相关概念等，以确定用户的真实意图；
- ③ 将用户的真实意图形成统一、规范的检索请求提交给检索匹配部分；
- ④ 在对用户意图进行分析和交互的基础上补充和完善本体库中的相关知识；
- ⑤ 对检索匹配部分返回的检索结果进行处理、合并后返回给用户。

(2) 知识源部分主要对知识源进行收集，并对收集的知识源根据本体库中的知识进行标注和分析，对从知识源中抽取的知识进行转换，然后再对本体库中的相关部分进行补充和完善，建立对应的索引信息，放入索引库。

(3) 检索匹配部分主要是从人-机交互部分收集统一的检索请求，并依据本体库中的相关知识对检索请求和索引库进行语义与语法层面的匹配，并将检索结果返回给人-机交互部分。

(4) 本体库部分应该说是整个模型的核心部分，从对检索请求到对检索请求与索引的匹配，再到对知识源的标注、索引的建立以及检索结果的处理都基于本体库中的相关知识。同时，上述各个过程又可以对本体库中的知识进行补充和完善。当然，对本体库中知识的任何修改都要经过领域专家和系统的双重认定。

7.3.4 分类检索模型

分类检索模型利用事物之间最本质的关系来组织资源对象，具有语义继承性，并可

揭示资源对象的等级关系、参照关系等，充分表达了用户的多维组合需求信息。

分类检索模型的核心思想是数据抽象，它利用事物之间最本质的关系来组织资源对象。其中，概念逻辑与知识分类思想是知识标引和知识检索的理论基础；知识组织结构是类层次结构；知识对象由静态和动态的知识元素构成，对象之间存在类属关系和其他语义相关关系。而且用户提问对象与知识库中知识对象的结构相同，可包含用户的多维异构需求信息。该模型综合应用类结构的查找方法和对象归类方法，实现了对对象或概念的检索。对于对象之间的等级关系，可提供快速的自顶向下的查找策略。利用继承原理，可实现自底向上的检索。其他语义相关关系，如参照关系，能实现横向的搜索。

这种模型的最大优点是语义继承特性，使得便于知识资源的共享，并可提供多途径、多方位的多种检索方法，充分表达了用户的多维组合需求信息。该模型较适于领域知识和用户知识等大规模知识源的情况。它需要综合利用自动化分类方法和手工分类方法，其大规模资源的存储和动态维护已由计算机实现。

7.3.5 多维认知检索模型

多维认知检索模型的理论基础是人工神经网络。它模拟人脑的结构，将信息资源组织为语义网络结构，利用学习机制和动态反馈技术，不断完善检索结果。

多维认知检索模型的语义网络由节点和链接组成。节点可以表示概念或知识对象，链接可以表示对象之间的各种语义关系和动态操作关系。该模型利用启发式知识和传递激活技术，形成一种认知式的语义推理模式，实现了知识库的联想搜索。它利用学习规则和相关反馈技术来修改、完善提问和知识库，不断改进检索结果，特别适合于专业概念知识库的检索，可用于扩充和优化用户的知识需求，还可帮助用户识别多义词的含义。用户可以充分描述概念与概念之间的关系，查找精确的文献内容和执行语义链的导航搜索。但它不适于大规模的知识源。

7.3.6 分布式检索模型

分布式检索模型综合利用多种技术，评价信息资源与用户需求的相关性，并在相关性高的知识库或数据库中执行检索，然后输出与用户需求相关、有效的检索结果。

对于网络异构分布式信息资源，分布式检索非常重要。该模型要求建立元知识仓库，集成和存储各类元知识。它综合应用分布式人工智能、神经网络、智能演算、并行推理和机器学习等技术，评估各类资源与用户需求的相关性，选择最好的知识源和数据库集合，分别执行之后并行检索。最后，它利用聚类、综合分析和学习等智能处理方法，产生全局一致的、有效的检索结果。元知识仓库（meta-knowledge repository）包含描述各类数据库内容的元数据、用户和专家的知识、各类控制知识、各种优化的分布式检索算

法和协作调控协议等。

7.4 检索结果可视化

知识检索系统不仅注重语义,而且更关注于检索结果的后处理,即如何把信息中隐含的知识利用某种推理模型以结构化的方式表示出来。检索结果的呈现目的是帮助用户从检索结果中抽取知识。

知识检索呈现要解决的问题就是知识提供的形式和检索结果显示的形式。知识按其结构来说可以分为层次型和网状型。依据已有的信息可视化相关研究,可以把知识的呈现分别表现为树结构和图结构。

知识检索结果可视化的作用主要体现在以下三个方面:

(1) 知识可视化有助于知识的传递。知识的传递是无处不在的,知识可视化可以提供一个系统的方法,来解决知识的传递并能提高知识传递的速度和质量。

(2) 在知识管理中,知识可视化为新的知识创新提供了很大的潜力。知识可视化使用了图像的创造力和图像的柔性布置的可能性(如图像可以根据用户的选择,逐步详细地显示用户需求的信息),从而创造新的知识。

(3) 知识可视化的应用是解决信息超载的有效策略。一般来说,信息超载是知识密集组织和信息社会中的主要问题。知识可视化通过分析框架、理论和模型能够压缩大量的信息。

7.4.1 呈现方法

知识检索结果呈现方法主要借鉴信息检索可视化模型,包括多参考点模型(MRPBM, The Multiple Reference Point Based Models)、欧几里得空间特征模型(ESCBM, The Euclidean Spatial Characteristics Based Models)、自组织图(SOM, The Self-organizing Maps Models)、路径搜寻相关网络模型(PFNET, The Path finder Associative Network Models)、多维尺度模型(MDS, The Multidimensional Scaling Models)和树地图(Tree Maps)。这些模型是目前可视化中比较成熟、主流且并被广泛应用和得到认可的模型,同时也是信息检索可视化中具有代表性的模型。上述每种模型都能够适用于多种情况,派生出一组相关的算法,有很大的适应性和扩展性。同时,这些模型能代表信息检索可视化的特点,揭示信息检索对象之间的深层语义和复杂的关系^[9]。

1. 多参考点模型

参考点(reference point)是一种信息检索的标准,利用这种标准可以从数据库里检索相关的信息,从广义上讲,它代表了用户的信息需求和任何与用户需求相关的信息。参考点可以是当前或者以前的检索词、检索出来的文献、用户的检索偏好和用户的背景知识,等等。参考点主要是通过提供一种体系保证检索的正确性,通过参考点可以辅助

修正原始查询，得到更加符合特定用户的查询结果。

参考点可以是固定的也可以是动态的。下面介绍一种基于固定多参考点的模型——InfoCrystal，这是一种针对布尔模型而设计的系统。Info Crystal 是用二维空间来可视化检索结果，参考点是查询式中的检索词，也可以是多个检索词的集合，这样就构成了多参考点。

图 7.7 显示了一个固定多参考点环境下的 4 个参考点构造模型。在图中， r_1 、 r_2 、 r_3 和 r_4 是分布在正方形四个角上的四个参考点，图中有四个层（tier）是因为有 4 个参考点。第一层里的结果表示只与一个参考点相关的文档，第二层里的结果表示与两个参考点相关的文档，以此类推，第三、四层里的结果分别表示与第三或第四个参考点相关的文档。例如， r_1 附近的圆表示与 r_1 相关的结果集， r_1 和 r_2 中间的长方形表示与 r_1 和 r_2 相关，每个层次的结果集用不同的图标表示，使人一目了然。这个模型中充分体现了布尔模型中的 AND 思想。

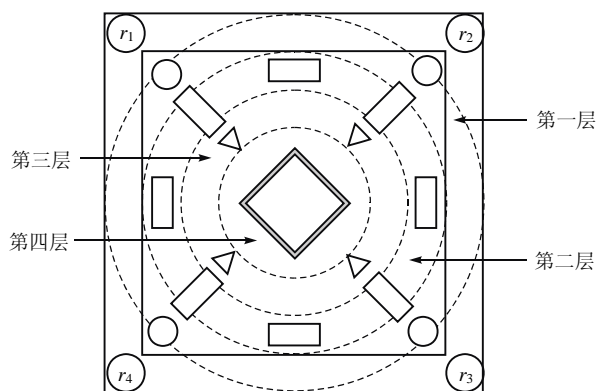


图 7.7 固定多参考点环境下的 4 个参考点演示图

2. 欧几里得空间特征模型

欧几里得空间特征主要体现在方向和距离上，这种方向性和距离性的特征可以应用在信息检索上。距离性是指在欧几里得空间中，距离越近的两个点越相似；方向性是计算欧几里得空间中两个向量之间的夹角，夹角越小则越相似，但这种相似是比例性的相似。在信息检索的余弦向量模型中，判断相似性的标准采用的就是方向性。因此在信息检索评价相似性时，如果是比例性的相似，就应该采用方向性的算法；如果是评价完全相似，则应该采用距离算法。

就欧几里得空间特征，下面介绍三种可视化模型：DARE、TOFIR 和 GUIDO。这三种模型都是在二维空间中可视化展示方向、距离或者两者的结合，一般用两个参考点来构造可视化空间，一个叫做 KVP (Key View Point)，另一个叫做 AVP (Auxiliary View Point)。

DARE 是一个基于距离——角度的模型，它展示了查询和检索到文档之间的语义关系，能够可视化余弦模型和距离模型，支持非传统的不对称的信息检索模型。在一个多

维空间中, 不管维数多高, 两点之间的距离和角度都是绝对存在的, 也是可以计算的, 这样以距离和角度来建立一个直角坐标, 可以将所有高维空间的文献映射到这个二维空间, 得到一个开放性的长方形的映射空间。用户可以在这个二维映射空间中划一条水平线, 将这条线沿 Y 轴移动, 就可以限制检出文献的数量。

TOFIR是一个基于角度——角度的模型, 它首先定义两个参考点 R_1 和 R_2 , 并分别定义为KVP和AVP, 则任何一个文档在高维的向量空间中与这两个参考点都有一个夹角, 将这两个夹角映射到二维空间中的两个坐标, 可以得出这个可视化空间是一个三角区域。有了这个三角区域的可视化空间之后, 就可以设置各种参数来确定检索区域。

GUIDO是一个基于距离——距离的模型, 它运用文档 D 与两个参考点之间的距离来映射, 映射后得到的可视化演示区域如图 7.8 所示。这是一个由三条线确定的开放性长方形区域, R_1 和 R_2 是两个参考点, 由于文档 D 到 R_1 和 R_2 的连线与 R_1 和 R_2 的连线构成了一个平面上的三角形, 因此得到了这个可视化区域。

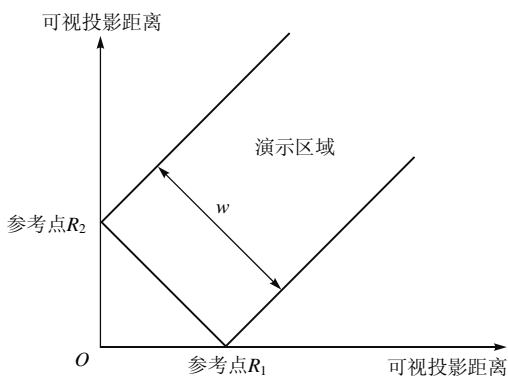


图 7.8 GUIDO 的可视化演示区域示意图

得到这个可视化的区域后, 就可以用各种模型来确定检索结果的区域, 如距离模型、椭圆模型和 Cassini 模型等。

在距离模型中, 如果两个参考点中的 R_1 被定义为KVP, R_2 被定义为AVP, 则 R_2 只是提供一种辅助的功能(如以前的查询式), 可以将高维空间中闭包上的点到 R_2 的距离确定为固定值 d , 这样映射到图 7.8 中的二维空间后, 就是划一条到 X 轴距离为 d 的水平线, 这样就确定了一个检索的区域。

在椭圆模型中, 设文档到 R_1 和 R_2 的距离是一个固定值, 这样文档就是在以 R_1 和 R_2 为圆心的椭圆边框上, 映射到图 7.8 之后就可以得到一个可移动的长方形检索结果区域。

通过这些可视化的模型方法, 可以将多维空间中看不见的闭包映射到可视的二维空间中, 使一些高维空间看不到的检索结果文件能够方便地显示在二维空间的检索区域中, 除了将文献之间的关系映射出来, 还将一些模型也可视化了。以前的信息检索模型相对于参考点来说是对称的检索模型, 系统在两个参考点形成相同大小的闭包, 但是查询词

对用户的重要性是不同的,这就要求系统在不同的参考点旁形成不同的闭包,在高维空间中要做到这一点是非常困难的,但是映射到低维空间之后,这种非对称检索方式就很容易实现了。

3. 自组织图

自组织是人工神经网络的一个分支。人类在对科学艺术探索的过程中,虽然积累了很多宝贵的知识,但到目前为止,还没有人弄清人的大脑是如何存储处理信息的,但科学家们没有中止对人类大脑的探索,仍然开发出很多模型来模拟人的神经网络。

自组织图(SOM, Self-Organizing Map)方法最初是在20世纪70年代由一位瑞典专家提出来的,后来经过许多专家的不断完善,最终Kohonen简化和优化了前人的成果,提出了一个更加实用、健壮算法,并命名为Kohonen自组织图算法。自组织图的可视化空间是一个二维的网状特征图,它可以应用在基于向量的信息检索模型中。SOM中有一个学习器,能够自动地对输入数据进行处理并生成特征图,且具有处理模糊数据和进行复杂计算的能力。

自组织图的体系中包括一个输入层和一个输出层,输入层的数据经过神经网络系统进行处理后,生成一个网状结构的二维特征图并进行输出。在网状结构图中,每一个网状单元(神经元)对应于一个加权向量,用于存储、记录在学习过程中所获得的知识 and 经验。

自组织图算法的过程如下:首先在训练过程开始之前,对特征图节点中的所有加权向量进行初始化,使接近于零的值被随机分配到特征图节点的加权向量中(该处理有利于最后得到合理化的结果);输入的信号经过一定的算法处理后由系统进行推理,对每一个信号在可视化的空间中找到一个最匹配的节点(winning node),然后对这个节点周围一定范围的其他节点按照相似性进行调整,距离越远调整得越少,越近则调整得越多;随着不断的调整,这个范围越来越小并逐渐接近该节点以至最后汇合,这样,整个训练过程也就结束了。经过对数据库中的数据进行训练后,最终产生了一个特征图(feature map),以可视化的方式展示知识的组织结构。

4. 路径搜寻相关网络模型

一个复杂的网络体系经过路径搜寻相关网络模型(PFNET)的处理之后可以产生一个最简单的网络,也就是说它只保留网络中两个节点的最短距离。PFNET主要是利用平面空间中三角形两边之和大于第三边的原理,它的一个主要参数,即路径长度,其值不能大于节点的个数。

PFNET可以应用于信息检索的很多方面。例如,网络资源都是以超链接组织起来的,但由于超链接的复杂性,人们很难清楚地了解这些网络资源之间的关系,通过PFNET的计算就可以将多余的、重复的网络节点去掉而保持最简洁的网络。这样在信息检索的

时候, 从一个点到另一个点查找的就是最短路径点; PFNET 也可以应用于文献之间的关联分析, 比如文献之间经过互引或同被引而产生的关联, 或者通过文献语义产生的关联等; 也有研究人员用 PFNET 对检索提问式进行研究, 根据提问式对用户进行分类, 形成一个网络。

在 PFNET 生成相关网络后, 可以用该网络来修改叙词表、提供辅助检索功能以及帮助了解数据库的整个结构等。

5. 多维尺度模型

多维尺度模型是一个比较成熟的算法, 它在信息可视化中主要用于计算数据集中的数据的相似性, 并用矩阵来表示, 然后以这个矩阵为输入数据, 将高维空间中的数据进行降维后映射到低维空间。用在多维尺度模型中的数据不受数据分布和数据形态的限制。

在可视化对象之间的语义距离时, 多维尺度模型根据高维空间中的距离对低维空间中的距离进行动态调整, 如高维空间中的距离近, 就在低维空间中把相应的距离调近点, 如果距离远, 就在低维空间中把距离调远点, 这样动态地进行调整, 直到高维空间中距离的方差减低维空间中距离的方差小于某一固定值。由此看来, 多维尺度模型就是用低维空间中的抽象距离来模拟高维空间的抽象距离的。

6. 树地图 (Tree Maps) 模型

树地图可视化技术是 Johnson 和 Shneiderman 于 1991 年首次提出的, 它属于支持浏览计算机中分层组织文件的一项典型技术。树地图是利用层次结构的元素填充反复分割的矩形区域的这样一个空间填充算法。它将一排排数据作为能整理、按尺寸分类和填充颜色的矩形组, 以便于图形化地揭示其内在的数据模式。这种算法方便终端用户认清本身并不明显的复杂的数据关系。

7.4.2 实现技术

在实现检索结果可视化过程中, 很多可视化软件都采用了 Java Applet 技术。通过这一技术实现检索结果可视化的过程是这样的:

(1) 数据预处理。对原始数据加工处理后, 可为后续的步骤使用做好准备。数据预处理包括抽词、聚类和数据转换等方式, 主要是为了抽取那些有检索意义的词, 去掉那些虚词、副词等一些与检索主题无关联的词语。对于抽取出来的词语要进一步加工, 使词语能准确表达原始数据的主题概念, 对那些与主题概念关系不大的词语要排除掉。

(2) 映射。映射是通过对词语处理的一些工具, 将大量看似无关联的词语按照一定的规则, 使它们相关, 这个过程针对不同的可视化软件有不同的要求, 映射步骤的目的就是将抽取出的主题词按照某种规则排列起来, 为可视化人-机接口提供合理、可用的数

据,便于后面的可视化工具利用这些数据,对这些数据作进一步的处理,最终为用户提供可视化的界面。

(3) 利用 Applet 绘制图形、图像。Applet 程序只接收传递给客户端的数据内容属性方面的信息,而不接收处理过的庞大数据集。它根据用户的检索需求,通过算法以及传递的参数信息去调用上一步中的数据,然后对这些数据进行筛选和过滤,选择匹配用户需求的数据,并保存数据对象之间的关系,以为后面的可视化显示打下基础。

(4) 显示。这步骤的作用是运用 Applet 小程序使数据对象以可视化的形式展现在用户面前。由于该可视化界面中包含了用户要搜索的对象,以及与该对象有联系的一系列对象,它们之间的关系用连线表示,用户每选择一个对象,都会有多个对象与之相联系,所以呈现给用户的是层次或网状的图形的集合。这个集合从多个角度说明了对象与对象之间的关系,有一对多和多对多的关系。



思考与讨论题

- (1) 信息检索的主要模型有哪些?各自的优、缺点是什么?
- (2) 简述知识检索与信息检索的区别。
- (3) 如何评价知识检索效果?有哪些指标可以利用?
- (4) 请列举检索结果可视化的呈现方式,并查找相关商业软件或开源工具。



参考文献

- [1] 刘挺,秦兵,张宇,等. 信息检索系统导论[M]. 北京:机械工业出版社,2008.
- [2] 杨建林,孙明军. 知识检索概念辨析[J]. 情报理论与实践,2006,29(4): 465-467.
- [3] 张玉峰,李敏,晏创业. 论知识检索与信息检索[J]. 中国图书馆学报,2003,5: 23-26.
- [4] 马文峰,杜小勇. 知识检索研究[J]. 情报理论与实践,2006,2: 25-27.
- [5] 张学福. 基于概念图的知识模型组织及协作机制研究——以 CmapTools 为例[J]. 现代图书情报技术,2005,11: 34-38.
- [6] 王兰成,曾琼. 基于本体的知识检索模型及呈现技术研究[J]. 图书情报工作,2009.
- [7] 陆敏,杨发毅,彭骏. 基于本体的知识组织和知识检索[J]. 现代情报,2009,29(1): 144-148.
- [8] 余肖生. 知识检索的可视化模型研究[J]. 情报科学,2008,26(8): 1240-1243.
- [9] 张进. 信息检索可视化的主流途径[J]. 图书情报知识,2008,125:24-27.
- [10] 周宁,张芳芳,余肖生. 可视化技术在知识管理领域的发展[J]. 图书情报工作,2006,50(11): 68-71.

第 8 章

数据挖掘和知识发现

内容提要

数据挖掘和知识发现是一门新兴的信息处理技术。数据挖掘的主要任务包括降维与特征提取、关联规则、分类与回归、聚类和异常检测等。文本挖掘和Web挖掘是数据挖掘与知识发现技术领域两个新的重要分支。文本挖掘是以半结构化或非结构化的自然语言文本为对象的数据挖掘技术，包括文本数据预处理、文本分类和聚类、文本关联分析等技术；Web挖掘是数据挖掘技术在Web数据上的应用，包括Web内容挖掘、Web结构挖掘和Web使用挖掘。

本章重点

- 数据挖掘
- 文本挖掘
- Web 挖掘



8.1 数据挖掘和知识发现简介

计算机技术和通信技术的迅猛发展将人类社会带入了信息时代。近些年来，数据库中存储的数据量急剧增大，大量的数据在给人们提供方便的同时也带来一系列问题。由于数据量过大，超出了人们掌握、理解数据的能力，因而给正确运用这些数据带来了困难。数据挖掘和知识发现是 20 世纪 90 年代兴起的一门信息处理技术，它是在数据和数据库急剧增长，远远超过人们对数据处理和理解能力的背景下产生的，也是数据库、统计学、机器学习、可视化与高性能计算技术等多学科发展融合的结果。

知识发现是指从大量数据中提取有效的、新颖的、潜在有用的、最终可被理解的模式的非平凡过程。数据挖掘是整个知识发现过程中的一个重要步骤，它运用一些算法从数据库中提取用户感兴趣的知识。由于数据挖掘对于知识发现的重要性，目前大多数知识发现的研究都集中在数据挖掘的算法和应用上，因此很多研究者往往对数据挖掘与知识发现不作严格区分，把两个名词等同使用。

数据挖掘涉及各种各样的算法来完成不同的任务。所有这些算法都试图为数据建立合适的模型，利用算法来分析数据，并确定与所分析数据的特征最符合的模型。一般来说，数据挖掘算法由模型、偏好和搜索三部分组成。算法的目的就是找到适合于数据的模型，但必须使用一些标准来进行模型选择。所有的算法都要使用搜索与优化技术对模型进行搜索。

如图 8.1 所示，数据挖掘模型在本质上可分为预测型模型和描述型模型两类。在图中可以看到，每类模型都用来完成一些数据挖掘任务。

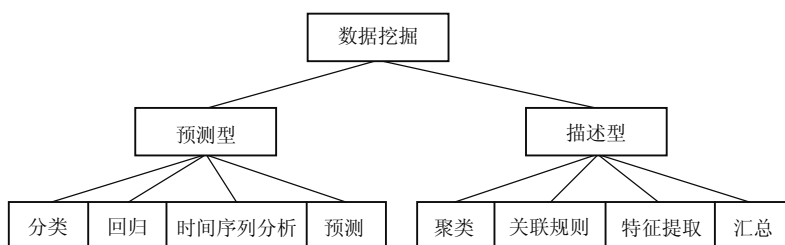


图 8.1 数据挖掘模型与任务

利用从不同数据中发现的已知结果，预测型模型对数据的值进行预测。它能够完成的数据挖掘任务包括分类、回归、时间序列分析和预测等。描述型模型对数据中的模式或关系进行辨识。与预测型模型不同的是，描述型模型提供了一种探索被分析数据性质的方法，而不是预测新的性质。聚类、关联规则、特征提取和汇总都通常被视为是描述型的。

8.2 知识发现的基本步骤

知识发现是一个包含了许多不同步骤的一个过程，这个过程输入的是原始数据，而过程的输出则是用户期望的有用信息和知识。然而，由于挖掘目标可能是不清楚或不准确的，因此过程本身是人-机交互的，而且可能要花费大量的时间。为了保证知识发现过程最终结果的有用性和准确性，整个过程都离不开交互作用，并且需要领域专家和技术专家的参与。图 8.2 列出了知识发现的全过程^[1, 2]。

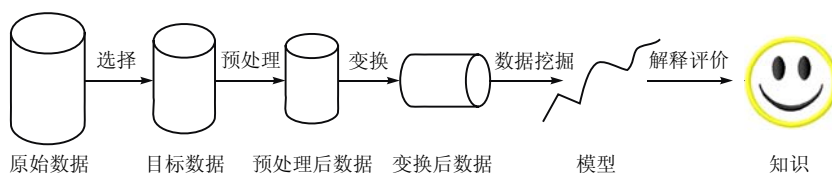


图 8.2 知识发现过程

整个知识发现过程由以下五个基本步骤组成^[1, 2]。

步骤 1：选择。数据挖掘过程所需要的数据可能从不同的异构数据源获取，因此，第一步就是从各种数据库、文件和非电子数据源中获取数据。

步骤 2：预处理。初始数据中可能会有一些错误的或者缺失的数据。由于数据源、数据类型以及度量的多样性，可能会有一些不规则数据，还会有一些同时实施的不同操作。错误的可以被修正或删除，但缺失的数据必须被补充或者预测（通常使用数据挖掘工具）。

步骤 3：变换。为了便于挖掘，从不同数据源获取的数据必须转换成统一的格式。一些数据可能需要编码或者变换成更容易使用的格式。可能需要采用数据约简来减少所考虑的数据属性值的数量。

步骤 4：数据挖掘。这一步骤就是基于所进行的数据挖掘任务，应用算法于变换后的数据来产生期望的挖掘结果。

步骤 5：解释/评价。数据挖掘的结果如何提交给用户是一个非常重要的问题，这是因为数据挖掘结果的有用性主要取决于这一步。在知识发现最后一步，通常使用各种可视化工具和图形用户界面来展现结果。

为了使数据适于挖掘和更加有用以及提供更有意义的结果，需要使用数据变换技术，即对数据的实际分布需要一定的变换以便使得要求特殊类型数据分布的技术更容易使用。在某些情况下将实值属性离散化可能更适于处理。有些数据可能还需要剔除，如异常点和不经常发生的极端值等。还可以利用函数对数据进行变换，如一个常用的变换函数就是对数函数，即使用数值的对数而不是数值本身。不论是降低维数（属性数），还是

减轻数据值的可变性，以上这些技术都使得数据挖掘任务更加容易进行。异常点的剔除实际上可以提高挖掘结果的质量。但在整个知识发现过程中，进行变换时一定要谨慎小心。如果错误地使用变换改变了数据，则数据挖掘的结果就会不准确。

这里所说的可视化是指数据的视觉展现。当考查数据结构的时候，可视化是非常有用的技术。例如，一个展现数据变量分布的折线图要比用公式表示的数据变量分布更容易理解，并可能提供更多的信息。将挖掘结果的数学符号表示与文本型描述相比，可视化技术使用户更容易概括、抽取和掌握复杂的结果。可视化工具不但可以作为一项数据挖掘技术来汇总数据，而且数据挖掘任务的复杂结果也可用可视化技术来展现。

数据挖掘过程本身也很复杂。有许多不同的数据挖掘应用方法和算法。为了使算法更加有效，每种算法都要谨慎地使用。而为了保证挖掘结果准确和有意义，一定要正确地解释和恰当地评价发现的模式^[3, 4]。

8.3 数据挖掘的主要任务和基本方法

数据挖掘的主要任务包括降维与特征提取、关联规则、分类与回归、聚类和异常检测等。

8.3.1 降维与特征提取

在数据挖掘中，一个经常碰到的情形就是数据具有高维特征。传统的数据库模式都是由许多不同属性组成的，但在求解某一给定的数据挖掘问题时可能并不需要全部属性。事实上，其中的一些属性可能会对数据挖掘任务的正确执行造成干扰，而另一些属性则可能增加算法的复杂性并降低算法的效率。这个问题有时被称作维数灾难，即由于涉及属性过多，导致难以确定使用哪些属性。高维问题的一个解决方案是降维，即减少属性的个数。但是，确定哪些属性是多余的，并非能够轻易完成。在模式分类、回归分析和聚类分析等数据挖掘任务中，降维通常是一个重要的数据预处理步骤。降维的主要方法包括特征选择和特征提取两类。

特征选择包括信息增益、互信息和卡方检验等多种方法，其中信息增益和卡方检验是比较常用的两种方法。特征选择方法的基本思想是首先将各个特征的重要程度进行量化，然后根据各个特征的重要性得分值进行选择。因此，如何量化特征的重要性，建立特征评估函数，就成了各种特征选择方法间最大的不同。以模式识别为例，在卡方检验中使用特征与类别间的关联性来进行这个量化，关联性越强，特征得分越高，该特征越应该被保留。在信息增益中，重要性的衡量标准就是看特征能够为分类系统带来多少信息，带来的信息越多，说明特征越重要。

特征提取是指对输入模式的原始观测数据所进行的一组变换,以便在比原始观测数据维数较低的特征空间对模式进行有效的描述或分类。在模式识别系统设计等实际问题中,原始数据经常包含一些多余的或重复的信息,为了减少整个识别系统获取原始观测数据的费用和相应的计算工作量以及改善识别系统的性能,有必要通过特征提取把模式变换到较低维数的特征空间中去。事实上,特征提取是模式识别的一个关键步骤。

设数据模式在测量空间中用 D 维向量 \mathbf{x} 表示,在特征空间中用 d ($d < D$) 维向量 \mathbf{y} 表示,线性特征提取器就是把 \mathbf{x} 变换为 \mathbf{y} 的 $d \times D$ 阶的一个矩阵 \mathbf{A} ,即 $\mathbf{y} = \mathbf{A}\mathbf{x}$ 。特征抽取的方法主要有主成分分析、独立成分分析、因子分析以及多维尺度分析等多元统计分析方法。

在实现一个具体的模式识别系统时,通过特征提取所求得特征向量,在很大程度上决定了识别系统的性能。因此一方面要对识别对象的各个方面进行深入的分析,尽可能从物理上确定某些对识别有效的特征;另一方面要与模式分类器的设计结合起来,反复进行试验,借以得到满意的结果。

8.3.2 关联规则

关联规则发现是数据挖掘中最重要的任务之一,它的目标是发现事务(Transactions)数据库中项目(Items)之间引人关注的关联。这些关联通常被用于零售业以了解哪些商品频繁地被顾客同时购买。

被用于寻找关联规则的事务数据库可以看作是一个元组集合,其中每个元组包含一组项目,如元组{面包、果冻、花生酱}就包含了3个项目:面包、果冻和花生酱。一个事务数据库中的关联规则挖掘可以用如下定义进行描述^[1]。

定义 8.1 给定一组项目 $I = \{I_1, I_2, \dots, I_m\}$ 和一个事务数据库 $D = \{t_1, t_2, \dots, t_n\}$, 其中 $t_i = \{I_{i1}, I_{i2}, \dots, I_{ik}\}$ 并且 $I_{ij} \in I$, 关联规则是形如 $X \Rightarrow Y$ 的蕴含式, 其中 $X, Y \subset I$ 是两个项目集合, 称为项目集, 并且 $X \cap Y = \emptyset$ 。

定义 8.2 关联规则 $X \Rightarrow Y$ 的支持度(s)是数据库中包含 $X \cup Y$ 的事务占库中所有事务的百分比。

定义 8.3 关联规则 $X \Rightarrow Y$ 的置信度或强度(α)是包含 $X \cup Y$ 的事务数与包含 X 的事务数的比值。

定义 8.4 给定一组项目 $I = \{I_1, I_2, \dots, I_m\}$ 和一个事务数据库 $D = \{t_1, t_2, \dots, t_n\}$, 其中 $t_i = \{I_{i1}, I_{i2}, \dots, I_{ik}\}$ 并且 $I_{ij} \in I$, 关联规则问题是找出满足最小支持度阈值和最小置信度阈值的所有关联规则 $X \Rightarrow Y$ 。数值(s, α)作为问题的输入而给定。

例 8.1 对表 8.1 中包含五个事务和五个项目{啤酒、面包、果冻、牛奶、花生酱}的超市购物篮事务数据库进行关联分析。

演示关联规则的样本数据如表 8.1 所示。

表 8.1 演示关联规则的样本数据

事 务	项 目
T1	面包、果冻、花生酱
T2	面包、花生酱
T3	面包、牛奶、花生酱
T4	啤酒、面包
T5	啤酒、牛奶

首先可以计算出整个项目集合的所有子集的支持度。如表 8.2 所示。

表 8.2 整个项目集合所有子集的支持度

项 目 集 合	支 持 度	项 目 集 合	支 持 度
啤酒	40%	啤酒、面包、牛奶	0%
面包	80%	啤酒、面包、花生酱	0%
果冻	20%	啤酒、果冻、牛奶	0%
牛奶	40%	啤酒、果冻、花生酱	0%
花生酱	60%	啤酒、牛奶、花生酱	0%
啤酒、面包	20%	面包、果冻、牛奶	0%
啤酒、果冻	0%	面包、果冻、花生酱	20%
啤酒、牛奶	20%	面包、牛奶、花生酱	20%
啤酒、花生酱	0%	果冻、牛奶、花生酱	0%
面包、果冻	20%	啤酒、面包、果冻、牛奶	0%
面包、牛奶	20%	啤酒、面包、果冻、花生酱	0%
面包、花生酱	60%	啤酒、面包、牛奶、花生酱	0%
果冻、牛奶	0%	啤酒、果冻、牛奶、花生酱	0%
果冻、花生酱	20%	面包、果冻、牛奶、花生酱	0%
牛奶、花生酱	20%	啤酒、面包、果冻、牛奶、花生酱	0%
啤酒、面包、花生酱	0%		

其次，按照不同的支持度和置信度以及关联规则的定义可以得到如表 8.3 所示的一些关联规则的支持度和置信度。

表 8.3 一些关联规则的支持度和置信度

关联规则 ($X \Rightarrow Y$)	支持度 (s)	置信度 (α)
面包 \Rightarrow 花生酱	60%	75%
花生酱 \Rightarrow 面包	60%	100%
啤酒 \Rightarrow 面包	20%	50%
花生酱 \Rightarrow 果冻	20%	33.3%
果冻 \Rightarrow 花生酱	20%	100%
果冻 \Rightarrow 牛奶	0%	0%

由关联规则问题的定义可知, 关联规则的选择取决于支持度和置信度。置信度度量规则的强度, 并支持度量规则在数据库中出现的频率。典型的情况使用较大的置信度数值和较小的支持度数值。例如, 考查表 8.3 中的规则: 面包 \Rightarrow 花生酱。 $\alpha = 75\%$ 表明这条规则在 75% 的情况下成立。也就是说, 在面包出现的情况下, 有 3/4 的时候花生酱也出现。与果冻 \Rightarrow 牛奶相比, 这是一条更强的规则, 因为在购买果冻时, 一次也没有购买牛奶。一个广告经理很可能不会根据一个顾客购买果冻时也购买了牛奶来开展广告活动。因为支持度表明了规则在整个数据库中出现的频率, 所以可以使用更低的支持度数值。例如, 果冻 \Rightarrow 花生酱的置信度是 100%, 但支持度只有 20%。可能这条关联规则只存在于 20% 的事务中, 但是当规则前件“果冻”出现时, 规则后件总是出现。这时针对购买果冻的顾客的广告策略是合适的。

Apriori 算法是最著名的关联规则算法, 已经为大部分商业软件所使用。该算法在关联规则挖掘过程主要包含两个阶段: 第一阶段必须先从数据集中找出所有的高频项目组 (Frequent Itemsets), 第二阶段再由这些高频项目组中产生关联规则。

关联规则挖掘的第一阶段必须从原始数据集中找出所有高频项目组。高频的意思是指相对于所有记录某一项目组出现的频率, 必须大于等于所设定的最小支持度阈值。一个满足最小支持度的 k -itemset, 则称为高频 k -项目组 (Frequent k -itemset)。算法从高频 k -项目组中再产生高频 $(k+1)$ -项目组, 直到无法再找到更长的高频项目组为止。

关联规则挖掘的第二阶段是要从高频项目组产生关联规则。利用前一步骤的高频 k -项目组来产生规则, 在最小置信度的条件阈值下, 若一规则所求得的置信度满足最小置信度, 称此规则即为关联规则。

从上面的介绍还可以看出, 关联规则挖掘通常比较适用于记录中的指标取离散值的情况。如果原始数据库中的指标值是取连续的数据, 则在关联规则挖掘之前应该进行适当的数据离散化。数据的离散化是数据挖掘前的预处理环节, 离散化的过程是否合理将直接影响关联规则的挖掘结果。

由于许多实际应用问题通常比超市购物篮分析问题更复杂, 因此已有大量研究从不同的角度对关联规则做了扩展, 将更多的因素集成到关联规则挖掘方法之中, 以此丰富关联规则的应用领域, 拓宽支持管理决策的范围。例如, 考虑属性之间的类别层次关系、时态关系和多表挖掘等。近年来围绕关联规则的研究主要集中于两个方面, 即扩展经典关联规则挖掘能够解决问题的范围和改善经典关联规则挖掘算法的效率。

8.3.3 分类与回归

分类是指将数据映射到预先定义好的群组或类。因为在分析观测数据之前, 类别就

已经被确定了, 所以分类通常被称作有指导的学习。分类算法通常通过观察已经知道所属类别的数据特征来描述类别。模式识别就是一类分类问题。输入的模式根据它与预先定义好的类别的相似度, 被划分到某一类中去。作为最常用的一种数据挖掘技术, 分类技术已经被广泛应用于许多领域, 如图像识别、医疗诊断、贷款审批、工业应用中的故障检测和金融市场走势分类等。分类与预测有着密切的关系。通常认为, 当被预测的值是连续值时, 称之为预测; 当被预测的值是离散值时, 称之为分类。回归分析是预测中最重要的一种建模方法。

进行分类的所有方法都假设已知训练数据的类别标签。通常利用训练集来计算分类技术需要确定的参数, 训练数据由样本输入数据与数据的类别归属组成。完成这个过程一般需要领域专家辅助。下面给出分类问题的定义^[1]。

定义 8.5 给定一个由元组 (条目, 记录) 组成的数据库 $D = \{t_1, t_2, \dots, t_n\}$ 和一个类别集合 $C = \{C_1, C_2, \dots, C_m\}$, 分类问题是指定义一个映射 $f: D \rightarrow C$, 其中每个元组 t_i 被分配到一个类中。一个类 C_j 精确地包含了被映射到它中的元组, 即 $C_j = \{t_i | f(t_i) = C_j, 1 \leq i \leq n, t_i \in D\}$ 。

上述定义将分类视为一个从数据库到类别集合的映射。注意到类别是预先定义的, 不重叠的, 且分割了整个数据库。数据库中的每个元组都被精确地分配到某个类中。对于一个分类问题的所有类别实际上是等价类。

实际上, 一个分类问题要分成如下两步来实现。

步骤 1: 通过对训练集进行计算产生一个特定的模型 (分类器)。在这个步骤中, 以训练数据 (包括对每个元组定义的类别) 作为输入数据, 以计算得到的模型作为输出数据。产生的模型要尽可能精确地分类训练数据。

步骤 2: 将第 1 步中产生的模型应用于目标数据库中的元组进行分类。

虽然实际上在步骤 2 才进行分类, 但更多的研究工作集中在步骤 1。步骤 2 通常是很简单的。有三种基本方法可以用来求解分类问题:

(1) 指定边界。这里是通过将潜在数据库元组的输入空间分割成区域来进行分类, 其中每个区域对应于一个类别。

(2) 利用概率分布。对于任意给定的类 C_j , $P(t_i | C_j)$ 是点 t_i 的类条件概率分布函数 (PDF)。如果各个类别 C_j 出现的概率 $P(C_j)$ 是已知的 (可能由领域专家确定), 则可用 $P(C_j)P(t_i | C_j)$ 来估计 t_i 属于类 C_j 的概率。

(3) 利用后验概率。给定一个数据值 t_i , 要确定 t_i 属于类 C_j 的概率, 记为 $P(C_j | t_i)$, 称作后验概率。分类方法就是要确定每一类的后验概率, 然后将 t_i 分配到具有最高概率的类中去。

根据分类算法所采用的基本思想, 可以给出不同类型的分类算法, 如图 8.3 所示。

其统计算法直接基于统计信息进行分类，基于距离的算法利用相似性或者距离度量进行分类，决策树、神经网络和支持向量机利用各自的结构进行分类，基于规则的分类算法则生成 IF-THEN 规则进行分类。

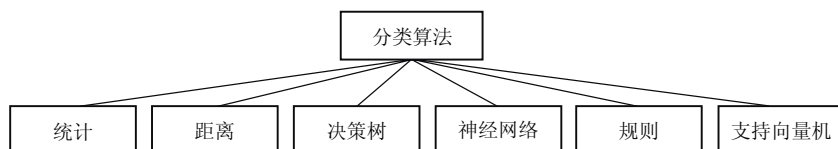


图 8.3 分类算法的分类

在实际的模式分类问题中，我们不但有多种不同类型的分类器可供选择，而且在每种类型的分类器中也有多种参数需要选择，比如在 k-最近邻分类器中选择多少近邻，在决策树分类器中的叶节点数目和在神经网络的隐层中有多少个神经元。不同的选择就对应着不同的分类器。一个自然的、判断分类器表现的评价准则是这个分类器做出错误分类的概率有多大。一个没有发生错误的分类器可能是完美的，但由于存在噪声数据，在实际中我们并不希望构建这样的分类器，这就是过拟合问题。如果分类器精确地拟合训练数据，则它可能不会很好地应用于更广泛的数据总体。解决过拟合问题的有效途径是依据结构风险最小化准则来选择和设计分类器。所谓的结构风险最小化就是在保证分类精度（经验风险）的同时，降低分类器的模型复杂度，从而使分类器在整个样本集上的期望风险得到控制。

回归是指将数据项映射到一个实值预测变量。事实上，回归涉及学习一个可以完成该映射的函数。回归首先假设一些已知类型的函数（如线性函数、Logistic 函数等）可以拟合目标数据，然后利用某种误差分析确定一个与目标数据拟合程度最好的函数。

下面用简单的一元线性回归公式描述一下回归建模过程。假设在训练样本中有 k 个点，则可以得到下列 k 个等式

$$y_i = c_0 + c_1 x_{1i} + \varepsilon_i \quad i = 1, 2, \dots, k \quad (8.1)$$

对于简单的线性回归，给定观测值 (x_{1i}, y_i) ，可以用平方误差技术来表示误差 ε_i 。为了使平方误差极小化，需要应用最小二乘法。用这种方法找到适当的系数 c_0 、 c_1 以使平方误差在观测值集合上最小化。误差的平方和为

$$L = \sum_{i=1}^k \varepsilon_i^2 = \sum_{i=1}^k (y_i - c_0 - c_1 x_{1i})^2 \quad (8.2)$$

取相应系数的偏导数并令其为零，求解后可以得到系数的最小二乘估计 \hat{c}_0 和 \hat{c}_1 。

回归分析与相关关系之间有紧密的联系。相关分析研究的是现象之间是否相关、相关的方向和密切程度，但一般不区别自变量或因变量。而回归分析则要分析现象之间相

关的具体形式，确定其数量依存关系，并用数学模型来表现具体关系。比如说，从相关分析中我们可以得知“质量”和“用户满意度”变量密切相关，但是这两个变量之间到底是哪个变量受哪个变量的影响，影响程度如何，则需要通过回归分析方法来确定。

一般来说，回归分析是通过规定因变量和自变量来确定变量之间的数量依存关系，建立回归模型，并根据实际观测数据来求解模型的各个参数，然后评价回归模型是否能够很好地拟合实测数据；如果能够很好地拟合，则可以根据自变量作进一步预测。

8.3.4 聚类 and 异常点检测

聚类作为数据挖掘的一个重要的研究领域，近年来备受关注。从机器学习的角度看，聚类是一种无监督的机器学习方法，它是将样本数据集合划分为由类似的样本点组成的多个类的过程。聚类方法作为一类非常重要的数据挖掘技术，其主要是依据样本点间相似性的度量标准将数据集自动分成几个群组，且使同一个群组内的样本点之间相似度尽量高，而属于不同群组的样本点之间相似度尽量低的一种方法。聚类中的组不是预先定义的，而是根据实际样本数据的特征，按照样本点之间的相似性来定义的，聚类中的组也称为簇。一个聚类分析系统的输入数据是一组样本和一个度量样本点间相似度（或距离）的标准，而输出则是簇集，即数据集的几个类，这些类构成一个分区或者分区结构。

假设簇的数目 k 是已知的，则容易给出聚类问题的定义^[1]。

定义 8.6 给定由一些元组组成的数据库 $D = \{t_1, t_2, \dots, t_n\}$ 和整数值 k ，则聚类问题就是定义一个映射 $f: D \rightarrow \{1, \dots, k\}$ ，其中第 i 个元组 t_i 被映射到第 j 个簇 K_j 中去。第 j 个簇 K_j 由所有被映射到该簇中的元组组成，即 $K_j = \{t_i | f(t_i) = K_j, 1 \leq j \leq k, t_i \in D\}$ 。

在上述定义中，每个簇 K_j ($1 \leq j \leq k$) 的实际内容（和解释）作为函数定义的结果而被确定。不失一般性，可将聚类问题的求解视为产生一个簇集 $K = \{K_1, K_2, \dots, K_k\}$ 的过程。聚类的过程可以分为特征选择和特征提取、聚类算法选择和设计、聚类验证以及结果解释和可视化四个基本步骤，如图 8.4 所示。

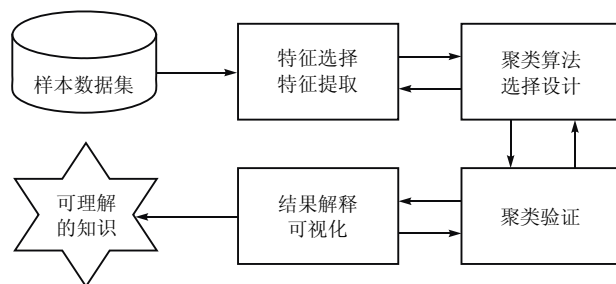


图 8.4 聚类的过程

根据聚类算法所采用的基本思想，可以给出不同类型聚类算法的分类，如图 8.5 所

示。其中层次聚方法与划分聚方法是最主要的两大类方法。所谓层次聚类是指产生一个嵌套的簇集。在层次体系中，每一层都有一些分开的簇；在最低层，每一个元组都组成一个单独的簇；在最高层，所有的元组都属于同一个簇。在层次聚类中，不必输入簇的数目。所谓划分聚类是指利用算法构造成一个簇集，其中簇的数目由用户指定或系统指定。传统的聚类算法为了满足内存要求，一般都是针对数值型的小型数据库设计的。但是，近来的许多算法都是针对大型动态数据库设计的，并且能够处理类别数据。为了满足内存约束，这些针对大型数据库设计的算法或者采取对数据进行抽样，或者利用数据结构来压缩或修剪数据库。基于是否产生重叠或非重叠的簇也可得到不同的聚类算法。

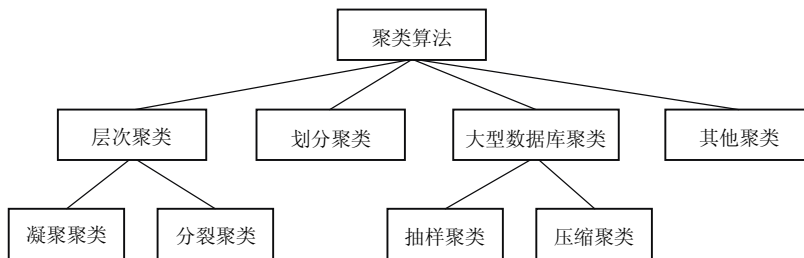


图 8.5 聚类算法的分类

基于具体的实现技术，可以进一步对聚类算法的类型进行分类。层次算法可分为凝聚聚类算法和分裂聚类算法两种。凝聚聚类算法是指产生簇的过程是“自底向上”，而分裂算法则是“自顶向下”。虽然层次算法和划分算法都可以用凝聚和分裂来描述，但这种描述更适于用来描述层次算法。根据算法对成员的处理方式不同，可分为串行（有时称增量）模式和同时处理模式。串行处理模式是指成员被算法一个一个地依次处理，而同时处理模式是指所有成员被算法一同处理。根据算法对成员属性值的处理方式不同，可分为单一处理模式和多处理模式。单一处理模式是指算法每次只考虑一个属性值；多处理模式是指算法一次考虑所有属性值。最后，按照算法的数学基础不同，聚类算法可以分为基于图论的算法和基于矩阵代数的算法。

聚类分析的一个附加的结果是对每个类的综合描述，这种结果对于更进一步深入分析数据集的特性是尤其重要的。聚类方法尤其适合用来讨论样本间的相互关联从而对一个样本结构做一个初步的评价。数据挖掘中的聚类研究主要集中在针对海量数据的有效和实用的聚类方法上，聚类方法的可伸缩性、高维聚类分析、分类属性数据聚类、具有混合属性数据的聚类和复杂类型数据聚类等问题是目前数据挖掘研究人员最为感兴趣的。

在聚类过程中，异常点的处理是非常困难的。所谓异常点是指不属于任何簇的成员。虽然它们可以被视为孤立的簇，但是如果一个聚类算法试图发现较大的簇，则这些异常点就会被迫被归入到某些簇中。由于将两个簇合并到一起并且保留了异常点的簇，这可能会导致产生效果不好的簇。

为了保证聚类效果，聚类算法可以发现并剔除异常点。但在实际剔除异常点时一定要谨慎。例如，假设数据挖掘问题是水灾预报，与正常水位值相比，非常高的水位值很少发生，因此可视为异常点。但如果剔除异常点，则数据挖掘算法就不能有效地预报水灾，这是因为反映水灾即将发生的数据已经被剔除了。

异常点检测或异常点数据挖掘是指在数据集中标识出异常点的过程。发现异常点后，利用聚类或者其他数据挖掘算法可以剔除它们或者按不同方式处理。许多异常点检测是基于统计技术的。这通常假设数据集服从一个已知的分布，然后通过众所周知的不一致性检验来检测出异常点。但是由于现实世界数据集不一定服从简单的数据分布，所以这种方法对于真实数据是不适用的。另外，大多数统计检验都假设使用单属性数值，而现实世界中数据集中的数据都是多属性的。此时采用基于距离测度的检测技术可能是一条可行的途径。

聚类和异常点检测已经被广泛应用于许多领域，如生物学、药学、人类学、市场营销和经济学。聚类应用包括动植物分类、疾病分类、图像处理、模式识别和文本检索。例如，在商业方面，聚类分析可以帮助市场人员发现顾客群中所存在的不同特征的群组，并可以利用购买模式来描述这些具有不同特征的顾客组群。在生物学方面，聚类分析可以用来获取动物或植物所存在的层次结构，并根据基因功能对其进行分类以获得对人群中所固有的结构更深入的了解。聚类还可以从地球观测数据库中帮助识别具有相似的土地使用情况的区域，此外，还可以帮助分类识别互联网上的文档以便进行信息发现。同聚类分析技术一样，异常检测也具有广泛的应用，如电信欺诈和信用卡欺诈、贷款审批、消费者行为分析、气象预报、金融领域客户分类和网络入侵检测等方面。

8.4 文本挖掘

文本挖掘就是从大量的文档中发现隐含知识和模式的一种方法和工具，它从数据挖掘发展而来，但与传统的数据挖掘又有许多不同^[5]。

8.4.1 文本挖掘简介

文本挖掘的对象是文档；文档内容是人类所使用的自然语言，缺乏计算机可理解的语义。传统数据挖掘所处理的数据是结构化的，而文档大都是半结构或无结构的。所以，文本挖掘面临的首要问题是如何在计算机中合理地表示文本，使之既要包含足够的信息以反映文本的特征，又不至于过于复杂使算法无法处理。

对文本挖掘技术的理解可以用图 8.6 来说明。图 8.6 由三部分组成：底层是文本挖掘的基础领域，包括数据挖掘与机器学习、数理统计、自然语言处理；在此基础上是文本

挖掘的基本技术,有五大类,包括文本数据预处理、文本分类与聚类、文本关联分析、文本信息检索与抽取、文本自动摘要;在基本技术之上是两个主要应用领域,包括信息访问和知识发现,信息访问包括信息检索、信息浏览、信息过滤、信息报告,知识发现包括文本数据分析与文本数据分类、聚类与关联分析等。总之,这里把对文本数据的预处理、信息检索、信息抽取和自动摘要以及从文本中发现知识都看作是文本挖掘。

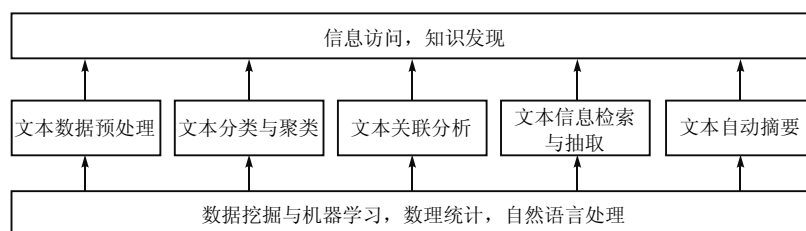


图 8.6 文本挖掘技术

对于人类而言,目前的大多数信息都是以文本数据的形式生成、呈现、存储、加工和处理的,而且随着互联网的快速发展,电子化的文本数量增长越来越快。同时,在企业、政府机构中,绝大多数的信息也都以文本的形式存在。因此,文本挖掘技术被认为具有较高的潜在商业价值。随着商务智能技术的广泛应用和商务智能市场的日趋成熟,以及文本挖掘技术的快速发展,基于文本挖掘的商务智能系统将得到长足的发展^[6,7]。

8.4.2 文本数据预处理

文本数据预处理是文本挖掘的一个基本问题,主要包括文本表示及其特征提取。文本表示是指将半结构或无结构的原始文本转化为结构化的计算机可以识别处理的信息的过程,即对文本进行科学的抽象,建立它的数学模型,用以描述和代替文本,使计算机能够通过对这种模型的计算和操作来实现对文本的识别。由于文本是非结构化的数据,要想从大量的文本中挖掘有用的信息就必须首先将文本转化为可处理的结构化形式。目前人们通常采用向量空间模型来描述文本向量,但是如果直接用分词算法和词频统计方法得到的特征项来表示文本向量中的各个维,则会产生一个高维向量空间。在高维空间中,这种未经处理的文本向量不仅给后续工作带来巨大的计算量,使整个处理过程的效率非常低下,而且会损害分类和聚类算法的精确性,从而使所得到的结果很难令人满意。因此,必须对文本向量做进一步处理,在保证原文含义的基础上,找出对文本特征类别最具代表性的文本特征。为了解决这个问题,最有效的办法就是对文本向量通过特征提取技术来降维。

目前有关文本表示的研究主要集中于文本表示模型的选择和特征项选择算法的选取上。用于表示文本的基本单位通常称为文本的特征或特征项。特征项必须具备以下特性:

① 特征项要能够确实标识文本内容；② 特征项具有将目标文本与其他文本相区分的能力；③ 特征项的个数不能太多；④ 特征项分离要比较容易实现。在中文文本中可以采用字、词或短语作为表示文本的特征项。相比较而言，词比字具有更强的表达能力，而词与短语相比，词的切分难度比短语的切分难度小得多。因此，目前大多数中文文本分类系统都采用词作为特征项，称作特征词。这些特征词作为文档的中间表示形式，用来实现文档与文档、文档与用户目标之间的相似度计算。

向量空间模型是文本挖掘中最常用的文本表示模型，本书在 7.1.3 节中已经做过介绍。向量空间模型把对文本内容的处理简化为向量空间中的向量运算，并且以空间上的相似度表达语义的相似度。

如果把所有的词都作为特征项，那么特征向量的维数将过于巨大，从而导致计算量太大，在这样的情况下，要完成文本分类几乎是不可能的。特征项的选择和提取应在保留文本核心信息的情况下尽量减少要处理的单词数，以此来降低向量空间维数，从而简化计算，提高文本处理的速度和效率。对于文本数据，除了可用 8.3.1 节中介绍的一般性方法进行降维外，还可用一些特殊的特征选择和提取方法。例如，停用词的过滤和关键词抽取等。文本特征提取对文本内容的过滤、分类、聚类、自动摘要以及用户兴趣模式发现等有关方面的研究都有非常重要的影响。

8.4.3 文本分类

文本分类是一种有指导机器学习问题，它需要事先定义一些主题类别。然后根据文本的内容自动将每篇文档归入其中的一个类别，这样用户即可以根据自己的所需来选择信息。从数学角度来看，文本分类其实就是一个映射的过程，它将未标明类别的文本映射到已有的一个或多个类别中。

与一般的模式分类过程一样，文本分类也分为训练和分类两个阶段，具体过程如下：

(1) 训练阶段，首先需要确定类别的集合 C ，这些类别可以是层次式的，也可以是并列式的。再选择适量具有代表性的文档组成训练文档集合 D ，确定训练文档集 D 中的每个训练文档 D_i 所属的类别 C_j ，然后提取训练文档 D_i 的特征，得到特征向量。通过对以特征向量表示的训练文档集进行计算产生一个特定的模型（分类器）。

(2) 分类阶段，将训练阶段产生的模型应用于测试文档集合 T 中的每个待分类文档进行分类。

文本分类本质上就是为一个文档分配一个或多个预定义类别的问题，因此文本分类的方法大部分来自于模式分类。例如，朴素贝叶斯分类、决策树、支持向量机、神经网络和 k -最近邻方法等。

8.4.4 文本聚类

文本聚类是指将文本根据其特征归类。也就是说,将给定的文本集合分为若干子集,称之为类,使得各个类内部的文本相似,而类与类之间的文本不相似。各种聚类方法原则上都可以用在文本聚类上。目前,有多种文本聚类算法,大致可以分为两种类型,即层次聚类法和划分聚类法。下面就层次聚类法(层次凝聚聚类法)给出文本聚类的简单过程描述。

对于给定的文档集合 $D=\{D_1, D_2, \dots, D_m\}$, 层次聚类法的具体过程如下。

步骤 1: 将 D 中的每个文档看作是一个具有单个成员的类 $C_i=\{D_i\}$, 这些类构成了 D 的一个聚类 $C=\{C_1, \dots, C_i, \dots, C_n\}$;

步骤 2: 计算 C 中每对类 (C_i, C_j) 之间的相似度 $\text{Sim}(C_i, C_j)$;

步骤 3: 选取具有最大相似度的类对,并将其合并为一个新的类,从而构成了 D 的一个新的聚类 $C=\{C_1, \dots, C_i, \dots, C_{n-1}\}$;

步骤 4: 重复上述步骤,直至 C 中剩下一个类为止。

层次聚类法构造出一棵生成树(在聚类分析中常被称为谱系图),包含了类的层次信息以及所有类内和类间的相似度,但是在每次合并时,需要全局地比较所有类间的相似度,并选择出最佳的两个类,因此运行速度较慢,不适合于大规模文档集合的聚类。

8.4.5 文本关联分析

文本关联分析是指从文档集合中找出不同特征词之间的关系。同文本分类与文本聚类一样,文本关联分析首先也要对文本数据进行预处理,然后才能调用关联规则挖掘算法进行关联分析。

在文档数据库中,可以将每个文档视为一个事务,文档中特征词或者关键词可视为事务中的项目,文档数据库可以表示为事务数据库的形式,即 $\{\text{Document_ID}, \text{Keyword_1}, \text{Keyword_2}, \dots, \text{Keyword_k}\}$,从而文档数据库中特征词关联挖掘的问题就可以映射为事务数据库中的关联规则挖掘。

值得一提的是,频繁地出现连续或者非常邻近的特征词可形成短语。文本关联分析与挖掘过程有助于找出领域相关的短语,即复合关联(Compound Associations),如[清华, 大学]、[美国, 总统, 奥巴马]和[有线, 电视]等;文本关联分析与挖掘过程也有助于找出领域相关的术语,即非复合关联(Noncompound Associations),如[人民币, 股票, 交易, 总额, 委托, 投资, 证券]。基于复合关联和非复合关联的挖掘统称为“术语级关联挖掘”。在文本分析中,术语级关联挖掘有两个优点:①可以自动标识术语和短语,

而无须人工标记；② 挖掘算法的运行时间很短且无意义的结果数量极大减少。在术语级关联分析与挖掘的基础上，还可以进一步进行文本分类和文本聚类等挖掘任务^[8,9]。

8.5 Web 挖掘

Web 挖掘是数据挖掘技术在 Web 数据上的应用，目前已经成为数据挖掘领域中的一个研究热点。

8.5.1 Web 挖掘简介

随着互联网在流量、规模飞速增长和复杂度增加等的形势下，万维网已经成为一个涉及诸如新闻、广告、消费信息、金融管理、远程教育、政府网站和电子商务等信息分布广泛的巨大知识宝库和信息海洋，而且万维网提供网络信息服务的竞争也日益激烈。

与此同时，一方面万维网上存储了大量的文档、图形和图像等类型的数据，表现出了 Web 数据的多样性；而另一方面，具有不同背景、兴趣和使用目的用户可以随意地链接到大部分的 Web 站点，Web 用户也呈现出多样性的特点。因此，万维网在给人们工作和生活带来极大便利和丰富信息资源的同时，也产生了一些亟待解决的问题。例如，难以准确获得所需要的信息，难以获得信息中潜藏的知识，欠缺个性化的信息服务等。要解决这些问题，必须根据 Web 数据的特性，对传统的数据挖掘方法进行扩展和改进，并结合统计学、计算机网络、数据库和数据仓库、可视化等众多领域的技术来对 Web 中蕴涵的多种信息进行挖掘。

一般地，从 Web 挖掘的对象可以将 Web 挖掘分为 Web 内容挖掘、Web 结构挖掘和 Web 使用挖掘^[10]三类，如图 8.7 所示。

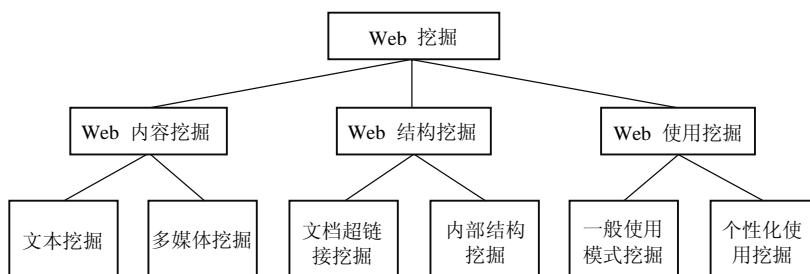


图 8.7 Web 挖掘分类图

在表 8.4 中对 Web 内容挖掘、Web 结构挖掘和 Web 日志挖掘中的数据特征、表现形式、挖掘方法以及应用领域等方面进行了总结和比较^[11]。

表 8.4 Web 内容挖掘、结构挖掘和使用挖掘的比较

	Web 挖掘			
	Web 内容挖掘		Web 结构挖掘	Web 使用挖掘
	信息检索领域	数据库领域		
数据	<ul style="list-style-type: none"> • 文本文档 • 超文本文档 	<ul style="list-style-type: none"> • 超文本文档 	<ul style="list-style-type: none"> • 链接结构 	<ul style="list-style-type: none"> • Web 服务器日志 • 代理服务器日志 • 浏览器日志
数据特征	<ul style="list-style-type: none"> • 非结构化 • 半结构化 	<ul style="list-style-type: none"> • 半结构化 • 视 Web 站点为一个数据库 	<ul style="list-style-type: none"> • 链接结构 	<ul style="list-style-type: none"> • 交互式数据
数据表示形式	<ul style="list-style-type: none"> • 无序/有序的特征词集合 • 术语和短语 • 概念/实体 • 关系曲线 	<ul style="list-style-type: none"> • 对象交换模型 (OEM) • 关系曲线 	<ul style="list-style-type: none"> • 图 	<ul style="list-style-type: none"> • 关系表 • 图
挖掘方法	<ul style="list-style-type: none"> • 向量空间模型 • 机器学习 • 统计方法 	<ul style="list-style-type: none"> • 信息抽取与集成 • 关联规则 	<ul style="list-style-type: none"> • PageRank 算法 • HITS 算法 	<ul style="list-style-type: none"> • 机器学习 • 统计 • 关联规则 • 聚类 • 序列模式
应用领域	<ul style="list-style-type: none"> • 分类 • 聚类 • 寻找抽取规则 • 寻找文本模式 • 用户建模 	<ul style="list-style-type: none"> • 发现频繁子结构 • 提取 Web 站点大纲 	<ul style="list-style-type: none"> • 分类 • 聚类 	<ul style="list-style-type: none"> • 站点结构管理及优化 • 网络销售 • 用户建模 • 推荐系统

8.5.2 Web 内容挖掘

Web 内容挖掘是指从 Web 上的网页内容及其描述信息中获取潜在的、有价值的知识模式，以实现 Web 资源的自动检索，提高 Web 数据的利用率。这项技术可以细分为 Web 文本挖掘和 Web 多媒体挖掘两种。Web 文本挖掘是对 Web 上大量文档集合的内容进行总结、分类、聚类和关联分析等。Web 多媒体挖掘是从 Web 多媒体数据如音频、视频和图像等中抽取事先未知的、隐藏的、完整的和新颖的知识。

Web 挖掘的另外一种分类体系把 Web 内容挖掘分为基于代理的挖掘和基于数据库的挖掘。基于代理的挖掘由软件系统（代理）负责内容挖掘，如搜索引擎就属于此类，它还包括智能搜索代理、信息过滤和个性化 Web 代理。与简单的搜索引擎相比，除了使用关键字之外，智能搜索代理还使用其他技术完成搜索任务。例如，可以使用用户信息和特定领域的知识。它的信息过滤使用信息检索技术、关于链接结构的知识和其他一些技

术实现对网页的存取和分类。个性化基于代理的挖掘以基于用户的偏好信息指导搜索过程；基于数据库的内容挖掘则把万维网上的数据看作数据库的数据，目前已经有一些方法把万维网看成一个多层次的数据库，并且已经存在多种面向万维网的查询语言。

事实上，可以将 Web 内容挖掘看作是对基本搜索引擎所完成工作的扩展。目前，已有许多基于最新的人工智能技术的搜索工具，这些工具能够从 Web 上更智能地提取信息，可以对 HTML 页面中的文本进行文本挖掘以及对页面中的多媒体信息进行挖掘。

8.5.3 Web 结构挖掘

Web 结构挖掘是指通过对 Web 站点的超链接结构进行分析、变形和归纳，将 Web 页面进行分类，以利于信息的搜索。Web 不仅由页面组成，而且还包含了从一个页面指向另一个页面的超链接信息。超链接信息包含了人类潜在的注释，大量的 Web 超链接信息提供了关于 Web 页面内容相关性、质量和结构方面的信息，它有助于自动推断出页面的权威性。当一个页面包含指向另一个页面的超链接时，可以认为是对另一个页面的认可。把对一个页面的不同超链接信息收集起来，就可以用来反映该页面的重要性。这就类似于信息检索领域中使用论文引用情况来评估该论文的质量。目前的主要方法是 PageRank 和 HITS。

PageRank 是一种由搜索引擎根据网页之间相互的超链接关系进行计算的网页排名技术，Google 用它来体现网页的相关性和重要性。因为 PageRank 通过网页之间的超链接关系来确定一个网页的等级。Google 把从 A 网页到 B 网页的链接解释为 A 网页给 B 网页的投票，从而就可以根据投票来源和投票目标的等级来决定当前网页的等级。简单的说，一个高等级的网页可以使其他低等级网页的等级提升。由 Google 的成功可以看出 PageRank 技术的有效性。

HITS (Hyperlink-induced topic search) 算法是另外一种使用广泛的链接分析算法，其基本思想也是利用页面之间的引用链接来挖掘隐含在其中的有用信息（如权威性），具有计算简单且效率高的特点。HITS 算法通过两个评价权值——内容权威度 (Authority) 和链接权威度 (Hub) 来对网页质量进行评估。HITS 算法包括两个步骤：① 基于一组给定的关键词（从查询中得到），发现一组（可能数千）相关网页；② 针对这些网页进行内容权威度度量 and 链接权威度度量，然后返回度量值最高的网页。基于 HITS 算法，IBM 公司开发了 Clever 网络信息检索系统。

8.5.4 Web 使用挖掘

Web 使用挖掘是指对用户和 Web 站点交互过程中留下的访问记录数据进行挖掘。这

些数据包括 Web 服务器的访问记录、代理服务器日志文件、浏览器日志记录、用户注册信息、用户对话或交易信息等。这些 Web 访问信息体现了用户使用 Web 资源的行为特性,以及隐藏在行为背后的更深层次的动因和规律。Web 访问信息的挖掘作为 Web 挖掘的一个重要组成部分,有其独特的理论和实践意义。其研究主要有两个方向:一般使用模式挖掘和个性化的使用特性挖掘。

一般使用模式挖掘主要是面向群体 Web 用户。通过 Web 使用挖掘,对总的用户访问行为、频度和内容等进行分析,可以得到关于群体用户访问行为和访问方式的普遍知识,用以改进 Web 服务方的设计。而更重要的是,通过对这些用户特征的理解和分析,可以有助于开展有针对性的电子商务活动,比如改进站点的组织结构以提供更高效的访问;吸引用户和保持用户;对用户的群体推荐;对用户总体地区、行业和阶层的分析等。

个性化的使用特性挖掘面向群体中的每一个用户。通过 Web 使用挖掘,对每个用户访问行为、频度和内容等进行分析,提取出每个用户或每类用户的特征,给每个用户或者用户簇提供个性化的电子商务服务。比如,个性化推荐、用户建模和个性化推销。个性化推荐是指如果用户的访问兴趣与其他一些用户很相似,那么就考虑推荐给该用户那些用户也感兴趣的一些东西;用户建模是指根据用户曾经的访问,推断当前访问的特征;个性化推销是指识别出对某种产品或服务的可能购买者,然后对其推荐相应的产品或服务。



思考与讨论题

- (1) 试述数据挖掘的主要任务与基本方法。
- (2) 什么是文本挖掘? 文本挖掘包括哪些主要内容?
- (3) 什么是 Web 挖掘? Web 挖掘包括哪些主要内容?



参考文献

- [1] 邓纳姆. 数据挖掘教程[M]. 郭崇慧, 田凤占, 靳晓明, 等译. 北京: 清华大学出版社, 2005.
- [2] Fayyad U M, Piatetsky-Shapiro G, Smyth P, et al. Advances in knowledge discovery and data mining[M]. California: AAAI press, 1996.
- [3] Han J, Kamber M. Data mining: concepts and techniques [M]. 2nd Edition, San Fransisco: Morgan Kaufmann Publishers, 2006.
- [4] 毛国君, 段立娟, 王实, 等. 数据挖掘原理与算法(第二版)[M]. 北京: 清华

大学出版社, 2007.

[5] 苗夺谦, 卫志华. 中文文本信息处理的原理与应用[M]. 北京: 清华大学出版社, 2007.

[6] 陈国青, 卫强. 商务智能原理与方法[M]. 北京: 电子工业出版社, 2009.

[7] 奥尔森, 石勇. 商业数据挖掘导论[M]. 吕巍, 等译. 北京: 机械工业出版社, 2007.

[8] 钱铁云, 王元珍, 冯小年. 结合类频率的关联中文文本分类[J]. 中文信息学报, 2004 18(6): 30-36.

[9] 宋擒豹, 沈钧毅. 基于关联规则的 Web 文档聚类算法[J]. 软件学报, 2002, 13(3): 417-423.

[10] 刘兵. Web 数据挖掘[M]. 俞勇, 薛贵荣, 韩定一, 译. 北京: 清华大学出版社, 2009.

[11] 苏新宁, 杨建林, 江念南, 等. 数据仓库和数据挖掘[M]. 北京: 清华大学出版社, 2006.

第 9 章

知识系统的技术体系结构

内容提要

知识系统的技术体系结构是指知识系统中的技术工具与方法的组成及相互间的关系。知识系统中的技术工具与方法在知识运作和管理中有着极其重要的作用。本章首先介绍知识系统中的人-机分工与协作，其次介绍信息网络、博客与维基、信息门户与知识门户、知识泵与个性化推荐系统、群件与计算机支持协同工作等技术工具与方法，最后介绍 Agent 技术及其在知识管理系统中的应用。

本章重点

- 信息网络在知识管理中的作用
- 信息门户与知识门户
- 知识泵与个性化推荐系统
- 群件与计算机支持协同工作



9.1 知识系统中的人-机分工与协作

本书在 3.4.1 节曾经说过,人类的各種知識是相互關聯的,形成了不斷發展演化的知識系統。知識系統中包含兩類子系統:一類是知識處理系統,另一類則是对知識處理進行管理的知識管理系統。這兩類系統形成了一個兩層結構。由于組織中的工作過程或人員也是以系統形式存在的,因此知識系統和業務工作過程系統或人員系統集成為一個更大的知識系統。

知識系統的 formed 和運行是以人和組織為主體的,但以計算機為中心的信息工具也起着重要作用。在考慮系統架構時,應該同時估計到系統中人的主動性與創造性以及計算機的能力與潛力,又要充分認識到兩者的局限性。表 9.1 列出了人与计算机在信息与知识处理等方面的比较。

表 9.1 人与计算机在信息与知识处理等方面的比较

	人	计 算 机
信息与知识处理的特征	主动、低速、模糊、并行	被动、高速、精确、串行
存储信息与知识的能力	低	高
擅长的信息与知识处理	决策	计算
记忆方式	联想记忆	寻址记忆
提高性能方法	学习	系统升级

虽然计算机在信息处理与知识处理的许多方面都具有超过人类的能力水平,但人类的创造性思维能力却是任何机器系统不可匹敌的。在创造性思维能力方面,机器系统的能力之所以不可能超过人类,一个重要的原因是作为机器系统设计者的人类自己,尚未完全理解创造性思维(如灵感、顿悟等)的机制,因此在设计机器系统的时候不可能为机器系统提供相关的指导性知识和原则^[1]。

由于人与机器各有所长,所以在知识系统中,人类发挥着主导作用,机器则是为了满足人和组织的需求在人主导下发挥辅助作用。尽管具体的以计算机与网络为核心的、支持知识管理的技术系统的构建与运行可以使用信息技术来完成,但是对于广义的知识系统来说,它的构建与运行则应该采用“人-机结合、各取所长、以人为本”的方式来完成。在构建人-机结合、以人为本的知识系统时,系统的结构既要有利于充分发挥信息工具的作用,又要充分提供人-机协同工作的可能性。

9.2 知识系统中的技术工具体系

知识系统中的技术工具与方法在知识运作和管理中有着极其重要的作用。

9.2.1 信息基础设施与计算机系统

人们在实现信息化进行信息管理与知识管理时，首先要构筑信息基础设施^[2,3]。

完整的信息基础设施包括：① 信息网络，包括企业内部和外部四通八达的信息网，可以传递数据、语音、图像和视频信息；② 信息装备，包括计算机硬件、软件和各种终端设备；③ 应用信息系统，即为企业生产、经营管理服务的各类应用系统，其中已嵌入有关的工作流程、操作规程以及相应的法规和制度等；④ 信息资源，即存储在各种媒体中的正式与非正式信息；⑤ 人员，即与信息系统有关的人员，既包括从事信息系统运用、提供信息技术服务的人员，又有使用信息系统来进行或协助自己业务活动的广大业务人员。

在知识管理领域中，使用了多种类型的计算机硬件和软件，由于知识管理使用信息工具是从信息管理沿袭而来，所以在硬件方面，从巨型机、大型主机到小型机、工作站、微型机都有所采用。从计算机系统的结构模式来看，主要有集中式系统、分布式系统以及集中与分布相结合的系统。

在企业信息化的初期使用的多是大、中型主机，以集中方式建立系统。所有的功能都集中在主机上。用户通过终端与主机进行沟通。在发达国家，这是 20 世纪 80 年代以前的主要结构模式。

但是知识运作与管理的大量出现是在 20 世纪的 90 年代，所以多数知识系统并没有建立在大型主机上，而是建在微型计算机的网络上。微型计算机的出现使得许多基层单位有条件为某些单一功能开发一些使用单台微机的知识应用系统，特别是中小型企业。后来多台单机连成网络，实现资源共享与相互通信，服务于一个共同目标，逐步发展出分布式计算环境（其中也可包含大型与小型机）。接着又出现了集中与分布相结合的系统，客户/服务器结构(C/S)模式，以及互联网上的浏览器/服务器/数据库(B/S/D)模式。

由于计算机性能和网络技术的迅速提高，今天的普通微机在性能上已与昔日的大、中型主机相当，所以近年来越来越多的系统采用了微机网络。网络成为 20 世纪 90 年代以来的主要结构模式。

使用计算机网络有下列好处：① 资源可以进行调剂；② 数据信息可以进行集中和综合处理；③ 均衡负载，相互协作；④ 系统可靠性可以提高；⑤ 系统易于扩展；⑥ 用户可以节约硬/软设备投资。

中国由于起步比发达国家晚，所以在使用主机的集中系统并不很多的阶段就进入了普遍使用微机阶段，接着就建立了大量网络系统，这可以说是一个跨越。

9.2.2 信息的类型

信息网络包括通信网、电视网、计算机网络几大类^[4]。现在的趋势是做到三类网络互连互通,形成统一的信息网络。当前计算机网络与通信网的连接已经有了很大的发展,这类相互连接的网络是知识系统所依托的信息网络。

这类信息网络的组成部分有:

(1) 网络节点,即用户进入网络的终端设备,如计算机等。由于信息技术的发展,许多移动设备,如手机,以及一些生产装置和业务处理设备也可作为节点进入网络。

(2) 网络连接介质,如电缆、光缆以及无线电连接等。

(3) 网络协议,网络运行时所应该遵循的实现规则 and 标准。

计算机网络按照其覆盖的地理范围不同而可分为:局域网(LAN)、城域网(MAN)和广域网(WAN)。

在一个企业或机关中使用的网络通常是局域网。它的出现比跨越广大地区的广域网要晚,但发展却快得多。它是企业或单位自建自用的,在一栋建筑或几百米的范围内。这类网络架设容易,传输速率高,其体验速率现在已经可以达到 1Gbps。

除了利用线缆组成局域网外,现在又出现了无线局域网,在一定的范围内(100~200m)通过无线电波传输,也可连成局域网。尽管它的传输速率目前还不够快,但它却提供了移动的方便,省掉了布线,这在建筑物已经装修完毕、很难再布线的情况,以及办公室的布置随时会改动、人员也会变动的情况下,显得十分方便。

小企业的信息系统通常就是建在一个局域网上。大企业的每个部门可以建立自己的局域网,再把各个局域网连接起来。现在的企业建筑物在施工时就采取综合布线把网线敷设好,这样就很容易组网和扩网。

城域网是在一个比较大的范围,如一个城市或者工业区内建立的信息网络是在 50km~80km 半径的范围。它基本上就是一个大型的局域网,通常使用的是局域网的技术。它可以覆盖一个地区的许多企业和单位。

广域网的覆盖范围很广,可以跨越城市、地区、省界,甚至于跨国、越洋。它是借用公用的通信网络连接起来的,其传输速率较低,一般应用在 64kbps~100Mbps,通过光纤传输。现在已经有 2.4Gbps 甚至更高速率的广域网。这种网络的拓扑结构复杂,管理也较困难。

广域网的重要组成部分是通信子网,它使用公用通信网来当作子网,如常用的公用电话交换网(PSTN)、数字数据网(DDN)、分组数据交换网(X.25)、帧中继(FR)、综合业务数据网(ISDN)等。

9.2.3 互联网与万维网

在所有的计算机信息网络中,规模最大、覆盖面最宽而且影响最大的,莫过于互联网(Internet),实际上它是国际性的广域网^[4]。从通信的角度来看,互联网是一个理想的信息交流媒体,能够迅速、高效地传递文字、声音、图像等各类信息,甚至通过这个网络召开远程视频会议。从获取信息与知识来看,它是一个世界上最大的信息和知识库。

互联网由专门的计算机网络设备如路由器等通过复杂的连接,把各地的广域网、局域网连成一体,以便于互相沟通。现在它已经把数以亿计的计算机连了起来。

互联网使用的是 TCP/IP 协议,它分成传输控制协议(TCP)和网络层协议(IP)两部分。每一用户有一个所谓 IP 地址。由于互联网采用的是数据分组(打包)传送,因而可以经路由器传送到网络的任何节点。当某一用户通过网络向任何一个其他用户发送数据时,TCP 协议先把数据分成若干个小的数据包,加上特定的标志,当数据到达目的地后,计算机去掉其中的 IP 地址信息,再利用 TCP 协议的装箱单检验数据有没有丢失(有丢失或损坏则重送),然后把数据包重新组合还原成原来的数据文件。互联网中有一种称为网关的设施,能够进行计算机本地语言与 TCP 网络语言之间的互相转换。

任何一台计算机,通过接入设备与互联网上的某一网络相连并进行登记,就可以与网上任何已登记的计算机或网络进行连接。接入设备可以是调制解调器,通过普通电话网(PSTN)或其他通信网(ISDN、ADSL 等)拨号接入,也可以通过与某一局域网连接而进入。

从功能上看,它原来能提供电子邮件、读取新闻、电子布告板(BBS)、远程文件传送、计算机远程登录等服务,但这些还都是以字符文本形式工作的。

现在一种新型的计算机模式——知识网格开始发展,这是利用互联网把分布在世界各地的计算机合成一台虚拟的计算机,使其具有超强的处理能力,并且充分利用网上限制的处理能力资源。在此基础上可以建立知识网格,使其成为互连的智能环境,辅助知识创新。

20 世纪 90 年代,欧洲核子物理研究中心(CERN)为了他们内部工作人员交流信息(不仅是文本,还有图形和图像等)方便,以互联网为基础研制出目前称为万维网(WWW, World Wide Web)这样一种应用系统,它现在已经是互联网中能力最强、使用最方便的一种工具了。

万维网以客户/服务器方式工作。在互联网中,有许多万维网的服务器,是信息的提供者,各服务器中存有它自身所提供服务的信息,同时还有其他服务器中信息的索引,这样在全球范围内各服务器相互指引,形成一个庞大的信息网。

在用户的计算机上运行的是万维网的客户程序,用户通过与互联网的接通,先访问

预先设定的服务器，查询有关信息，也可以通过它接通其他服务器，得到信息。

在万维网中，采用了图形接口、超文本、多媒体等技术。万维网使用的是超文本传输协议（HTTP），所谓超文本（Hypertext），是一种计算机文本组织与显示的技术，它事先对文本中有关的词汇进行索引链接，使其能指向文本中其他段落、注解或其他文本中的内容，用户通过鼠标对这些词语操作，便可显示该词语链接的内容，这样读取文本内容不是像过去那样逐页逐行逐字地线性操作，而是一种非线性的操作。使用多媒体技术，超文本又可演化为超媒体（Hypermedia），它可以使文字词语、图形、图像、视像、声音等相互链接，进行检索，大大扩展了人-机交互过程中计算机显示和检索的能力，因为它不需要知道信息在网络中的具体地址（IP 地址），也不必知道文件名称，只要按照超媒体的链接去一步步查找就行了，这就使得许多不掌握计算机专门技术的人也能使用。

在万维网中，信息是以页为单位存放的，页内和页与页之间的信息操作通过上面讲的超媒体、超文本方式。网页的编写使用的是超文本置标语言（HTML），用它编写的文档及文档内容，又包含文档的标记（文档格式与显示方式）。使用 HTML 的编辑器，很容易编写网页。还可以应用 JAVA 语言与技术，加入各种动态效果。现在一种新的、叫做可扩展置标语言（XML）也开始使用。它可以描述任何逻辑关系的数据格式，以统一电子数据的格式进行存储，从而不再因为数据格式的不统一而苦恼和困惑。

在万维网的应用中，用户不仅需要进行浏览，而且需要与网站进行交互操作。现在已经有了动态服务器主页（ASP）技术和专门主页（PHP）技术，使用户可以通过浏览器进行交互操作。在万维网上，信息的链接是通过三条并行的路径进行的，一条是按专题性质，一条是按地理位置，再一条是按业务性质（如文件传送和远程登录等）。

目前，客户程序有两大类。一类是字符操作型的，应用不太广泛；另一类是多媒体型的，可在各种小型机、工作站和微机上运行，这类程序应用很广，如当前流行的 Internet Explorer 以及 Opera、Firefox 等。用户运行这类称为浏览器的程序，一下子便在屏幕上看到所连的服务器上的主页，立刻可用鼠标选择所需信息。

这些浏览器把电子邮件（E-mail）、文件传送（FTP）等功能集成进去以后，成为非常通用的网络工具，对促进互联网用户的大量接入起了很大的作用。在知识运作与管理中，万维网是非常重要的知识来源和知识传递工具。

现在万维网上有一种新的运行方式 Web 2.0 开始流行。Web 2.0 并不是一款新的万维网应用软件，也不是新版的万维网技术标准，而是一种新的服务模式，一个新的视角。在 Web 2.0 的视角下，万维网不再是一些相互连接的站点的集合，而是为网民提供各种应用的平台。这些应用可以为用户的参与和彼此的协作提供极大的便利。本章在后面将要提到的作为知识资源的博客与维基，就都可以称为 Web 2.0 技术。

9.2.4 内联网与外联网

使用互联网技术而应用于企业内部的内联网（Intranet），是一种新的网络。它使企业网络的构建更加方便，还能通过互联网与外界联成一体，大大扩展了系统的范围和功能。用同样技术发展出来为企业间沟通服务的外联网（Extranet），则进一步扩大了企业与外界的沟通范围。

内联网是以互联网的技术、特别是其中的万维网（WWW）技术为基础而建立的，主要用于企业或组织内部的信息交流，但仍能与互联网连接的一种新型网络。从它们的英文名称看，只有个别字母的差别，Intra 指组织内部，Inter 则指组织相互之间，这正反映了它们的功能特点。前面已经提到，万维网（WWW）本身作为一种应用系统，具有统一的人-机交互接口和良好的信息表达方式，高度的用户友好性，方便的检索浏览功能，联系着全球可以公开检索的信息资源，因此自然而然地逐渐成为国际互联网的主体。尤其是许多公司、院所在它上面建立了自己的主页，因而为发展电子商务、科技交流提供了极为方便的手段。现在，发达国家的大中小企业几乎都与它连网，甚至一个大企业分散在各地的分支机构也通过它来传递信息。

在企业内部的、由自己的万维网服务器与员工的客户机形成的系统，主要是为企业内部的信息交流服务，但仍能通过国际互联网与全球各地相通，至于企业内部不宜向外交流的信息，则可采用防火墙技术来加以隔离。这样就自然而然地形成一种新的网络，这就是内联网。

正式提出内联网的名称是 1995 年以后的事，但一经提出，便有许多企业宣称它们已经建立了这种网络。任何一种新技术的产生都要受到需求牵动（或称用户导向）和技术推动（或称厂商导向）两方面的影响，现在有些信息技术和产品的产生与推广是受了厂商导向的影响、由技术推动的，但内联网的产生却是用户导向、需求牵动的，一旦出现了内联网这种称呼，便发现原来已经有了不少类似的系统。这时厂商便又集中力量去推动它的发展，使内联网成为近年来发展最快的组织内部的网络。

有人给内联网下的定义是：内联网乃是以互联网技术为基础的企业网络基础设施，它把企业系统、互联网、工作群体结合在一起。

内联网具有如下主要特点：① 它用统一的方式方法表示信息；② 它可以把企业或组织分布在各地的计算机、软件、数据库连成一个系统、使工作人员很容易找到所需要的信息，而不管它存放在什么地方；③ 它的协议和技术标准是统一而且公开的，可以同时支持多种机型和操作系统，成为跨平台沟通的基础；④ 目前它的客户软件可以从国际互联网上免费取用，这给网络建设节约了大量经费和时间；⑤ 它可以通过局域网以至于分组交换网、数字数据网等方式组网，非常灵活；⑥ 它既可用于企业内部信息交流（这

是主要的), 又可与国际互联网接通, 与国内外的组织机构、个人进行信息交流。内部信息又可加以隔离而不致外泄。

利用内联网代替传统的局域网有许多优点: 内联网具有很好的开放性, 很容易与互联网连接, 但又能通过防火墙进行安全隔离; 它的可扩展性很好, 更新也比较容易; 延伸范围比较大; 容易共享资源; 可以实现跨平台集成。特别是使用浏览器—万维网—数据库(B—W—D)工作模式, 可以把各种应用集中在服务器上, 用户只需要具有普通的浏览器就可以进行各种应用, 这就大大扩展了用户范围。

在企业内部互联网的基础上, 可以建立一种综合的现代企业信息系统, 从信息功能来看, 这类系统由下列四大平台技术构成。

(1) 网络应用支撑平台。这里使用广域网与局域网中已经广泛采用的 TCP / IP 协议, 结合广域网的互联、路由、网络管理、防火墙、虚拟专用网等技术, 建立一个安全稳固的开放网络应用平台, 支撑上层应用程序的运行, 是整个企业管理信息综合环境的基础, 也是重要的通信基础设施。

(2) 信息资源管理平台。它以万维网技术为基础, 应用 HTML 超文本多媒体开放文档体系与信息链, 建立虚拟的资料库, 实现信息的快速查询和获取。

(3) 消息传递与 workflow 管理平台。具有先进的消息传递与分布式目标管理, 追踪工作流程的用户化事务处理管理, 以及安全可靠的数据签名、身份验证和加密功能, 可以发布信息并及时掌握信息具体流向和进行反馈, 提高工作效率, 其中集成了电子邮件、个人与群组工作表、电子表格以及其他共享信息的应用系统。

(4) 业务处理平台。它是采用分布式处理结构和先进的数据库管理系统, 建立具有各种分析、预测、决策支持等功能的业务处理平台, 并可与外界互联网连通。

上面这些平台的大部分在知识系统中都是重要的组成部件。

外联网是一种利用互联网技术把本组织与其他组织或个人连接起来的网络, 它可以用来把公司与客户、总公司与各地的分公司、本公司与其他公司连接起来。它只是把有关的单位连接起来, 对公众是不公开的, 或者是有选择地公开的。它把企业内部网络扩展到企业之外, 使得企业与供应商、客户以及合作伙伴之间易于进行信息和知识的沟通。还可以用来进行如远程培训等异地活动。

外联网既解决了开放性与共享性问题, 又解决了安全性问题。对一个业经授权的集体成员来说, 能够非常方便地获取和交流信息与知识, 又不至于被未经授权的成员获取有一定私密性的信息和知识。

外联网可以通过下面几种方式来实现。

(1) 通过公共网络(如互联网)进行内联网的互连, 这是最方便的实现方法, 但是其安全性不能保证。尽管可以用防火墙来加以隔离以做到内外有别, 但防火墙不是绝对

安全的。所以这种方式较少采用。

(2) 通过专有网络来进行连接, 专有网络利用的是租用的电信网络的专线。它与一般拨号电话不同之处在于它是始终连通着的, 而普通电话只是在拨通之后才连通。它的突出优点是安全, 但成本太高, 尤其是一个组织要与多个单位连通时, 需要租用多条专线, 费用高而利用率低。

(3) 通过虚拟专用网络 (VPN) 实现。这是在公用网络上通过一定的协议建立通道和数据封装, 构建起虚拟的网络。它的优点是既保证了安全性, 又可降低成本。同时对公用网络来说也提高了线路利用率。

9.3 博客与维基在知识管理中的作用

博客与维基是两种新兴的知识管理工具, 具有广泛的应用前景。

9.3.1 博客与知识管理

博客的英文名称为 **Blog**, 是 **Weblog** 的简称。博客是网络上的一种信息组织记录形式, 所以也称为网络日志或网络日记。作为一种表达个人思想, 内容按照时间顺序排列, 并且能够自动更新的网络出版交流方式, 博客比电子邮件、讨论群组更加简单和方便易用, 并因而成为了继电子邮件、电子布告板 (BBS)、即时信息传递软件 (ICQ) 之后第四种颇受网络用户喜爱的网络信息组织和交流方式。

尽管博客的类型多种多样, 但都拥有大致相同的特征, 概括起来有以下几点: ① 个人性。写博客纯粹是个人的自发行为。② 即时性。写博客是一种习惯, 经常更新、不断积累是博客文体有别于其他个人文章、著作的关键。③ 开放性。博客是要把自己最珍贵、最有价值的收获都奉献出来。④ 网络连接。博客的优势就是通过网络连接等方式, 不断搜索提炼信息, 使之不断学习和思考。⑤ 交互性。博客与读者的交流是关键, 没有互动和交流的博客是缺乏活力的。⑥ 可信度。博客的可信度来源于竞争, 不可信的博客是很少有人访问的, 也就存在不下去。

博客作为一种新兴的网络交流工具, 被人们越来越广泛地应用于个人和组织的个性化知识管理之中。在知识管理过程中, 沟通至关重要。博客作为个人和组织知识管理中的一种有效沟通工具, 在知识过滤、知识转化、知识共享 and 知识应用四个方面都能够发挥重要作用^[5]。

(1) 博客的知识过滤功能。博客提供的内容只是各种各样的信息, 不是所有的信息都能成为知识, 它需要一个信息的收集和过滤的过程, 相关的隐性知识, 也要通过文档进行显性化处理, 以变得更容易为人所理解和接受。博客主要通过搜索引擎、超文本链

接和 RSS (Really Simple Syndication) 信息聚合器实现知识的积累。RSS 正成为博客的基本特征, 这种技术提供了分布式的信息聚合方式, 允许博客自动将更新的内容发送给接收者, 接收者可以自主选择信息来源, 自动获得来自其他博客上的信息, 并且可以通过超文本链接等方式对信息加工整理, 转化为自己的知识。

(2) 博客的知识转化功能。博客的零技术特点为众多不懂网络技术的人提供了参与的机会, 而零成本则吸引了更多的免费用户, 极大地扩展了用户群。从知识管理的角度来看, 用户可以通过博客查找并且精选各种资料内容, 同时又可与他人就某一主题进行持久深入、广泛的交流讨论, 因此基于博客的知识管理使学习者更容易实现兼具广度和深度的知识转化。此外, 博客对知识收集过程中获得的知识, 还可进行有意识地筛选和过滤。这个过程, 是从显性知识到显性知识的转化过程, 由此形成的知识更加系统和全面, 知识应用的可操作性更强; 同时这个过程, 也是由显性知识到隐性知识的转化过程, 通过它使自己头脑中所提取的知识进行过滤消化, 去伪存真, 以提高自己的认识深度。

(3) 博客的知识共享功能。博客用户经过知识收集和知识提取后, 能够获得知识上的提高, 在此基础上产生的心得体会, 又可以通过博客文章的形式, 传递到自己的博客上, 完成从隐性知识到显性知识的转化。由于博客是网络中的一个节点, 博客文章中一般包括关键字和摘要等, 这为人们利用搜索引擎进行知识共享提供了方便。其他人获得这些知识后, 可以进行相互评论和引用, 好像是召开网络会议, 每个人都可以就所关心的问题各抒己见, 从而形成广泛的知识共享。

(4) 博客的知识应用功能。博客用户将获得的知识, 应用于具体实践工作中, 可以发挥知识的实际作用, 通过实际工作, 来验证自己所获得的知识的正确性, 完善和充实自己的知识内容; 也通过自己的身体力行, 对其他人形成从隐性知识到隐性知识的传递。博客作为优秀的网络沟通工具, 在知识应用中发挥着作用, 通过博客反映应用知识的结果和成效, 即时传递实际工作中的背景信息, 获得其他人的看法、意见和解决问题的方法、思路, 通过参考他人的知识, 从而对自己的知识应用形成有益的帮助。

现在有些企业已经开始尝试采用博客技术进行知识管理。企业通过建立内部的博客社区, 为每个员工建立自己的博客, 提倡员工在工作中及时记录自己的心得、体会, 并通过阅读彼此的博客达到隐性知识交流的目的。博客可以实现隐性知识与显性知识之间的相互转化, 从而有助于促成企业知识的形成, 提高企业的知识管理能力。

9.3.2 维基与知识管理

维基 (Wiki) 一词源于夏威夷语 “Wee Kee Wee Kee”, 意思是 “快点快点”。维基是一种服务软件, 它允许用户通过 Web 搜索引擎, 自由创建和编辑 Web 网页内容。用户可以使用十分简单的、基于内容的标识, 在 Web 基础上对维基文本进行浏览、创建和

更改，而且创建更改发布的代价远比HTML文本要小。维基不仅给人们提供了一种面向社群的协作式写作，同时还提供一组支持这种写作的辅助工具。可以说，维基的目的就在于构建知识库，并且所有的人都可以参与知识库的构建，贡献自己的知识。因此，在Web环境下，维基使人们容易创造、组织和获得与其他网站（页）相链接的一个Web网站。这样，它就可能成为一个全社会信息交流的软件，通过它给人们提供一个一起工作和交流的机会^[6]。

由于维基编辑方便、标记简单，因此它使用范围较广，已成为一种新的互动开放的信息交流平台。维基的特点归纳起来有以下几方面：① 开放。社群的成员都可以创建、修改和删除页面。② 协作。维基是一个协作共创系统，要求社群的成员具有合作精神。③ 平等。社群中的成员是平等的，每一个成员都有相同的责任和权利。④ 共享。系统内页面的变动都可以被访问者以及社群的每一个成员所共享。

维基和博客都是网上信息新的发布、交流和展示方式，可以大大简化信息发布的复杂程度，改变知识旧有的传播模式，但两者区别较为明显^[7]，如表 9.2 所示。

表 9.2 维基与博客的区别

区 别	维 基	博 客
构成框架	通常由一系列随时可以再编辑的条目构成，并按照主题进行分类和组织	通常由一系列经常更新的流水纪录构成，并按照日期倒序排列
参与方式	多数人写、多数人看，每个人既是阅读者，同时又是书写者，相当于人人当记者和编辑	可以一个人或少数人写、多数人看，阅读者被动接受，只能在文后添加评论，相当于人人当记者
特征描述	以主题为主线，强调协作、客观和中立，适合社区的知识管理与知识积累，一般适合于创建用作资料库、知识库或信息库等。注重协作性，主题相对明确，内容关联性很强，组织紧密	以个人思想为主线，强调以个人为中心，适合个人的知识管理和言论表达，主题相对松散，基本不会刻意控制内容的相关性
精神体现	开放、协作、平等和共享	私人性与公共性的结合，具有一定的共享精神和价值
存储方式	动态存储，内容若删除仍可找到原文件	静态存储，内容若删除就很难再恢复

维基为企业知识管理提供了一种新的工具。维基系统所具有的开放、协作、平等、共享等特征和精神的体现，创造了人人都可以实现自我价值的平台，员工可以借助该平台充分表达自己的思想、意见和建议，为企业的正确决策提供广泛和坚实的群众基础。维基系统在企业经营管理中的应用还同时体现出企业对员工的信任程度，并由此影响到企业的经营成本。维基作为一种协作软件，它能集中和管理企业的各类数据、信息和知识。如摩托罗拉公司的维基系统可以支持员工在实践中进行知识交流、积累、共享和创新，促进隐性知识与显性知识的相互转化，创造各类新知识，帮助企业实现整体知识规模的拓展以及知识质量的提升。因此，应用维基进行知识管理是提高企业本身素质，提

升企业核心竞争力的有效措施之一。

9.4 信息门户与知识门户

企业信息门户和知识门户主要用于将分散的信息资源和知识资源整合到一起，以达到有效运用这些资源，进而提高工作效率和决策能力的目的。

9.4.1 信息门户

随着信息化建设的不断发展，企业内的信息系统的应用越来越广泛，它们为了适应业务的发展需要逐步建立了不同的应用系统，如企业资源计划（ERP）、财务管理信息系统（FMIS）、客户关系管理（CRM）、供应链管理（SCM）、电子化人力资源管理（e-HR）、办公自动化（OA）等管理系统，但由于历史原因，不同的应用系统之间缺乏整体的考虑，在应用过程中常会面临以下问题：① 系统间彼此孤立，数据分散；② 用户由于需要，必须频繁地登录不同的系统，使用方法不统一，在安全性方面缺乏统一的管理机制与控制；③ 对于企业内部不同工作岗位的人员来说，所需要的资源存放在不同的系统当中，导致频繁的在系统间来回切换；④ 同一信息在财务、ERP、CRM 等系统都可能保存，导致信息的不一致，不同的信息由于分散存储在不同的系统，也很难形成组织统一的信息模型；⑤ 当企业面临端到端的服务挑战时，却由于不同的业务流程在不同的系统流转而无法满足，企业的协同效率明显受到制约。这些现象已渐渐开始阻碍企业的管理和发展。

企业信息门户（EIP, Enterprise Information Portal）就是要针对上述问题而构建企业的统一门户，帮助企业实现以下目标：① 员工、客户、供应商和合作伙伴可以在任何时间、任何地点迅速获得个性化的内容、应用和服务，并有统一的访问方式；② 及时、准确地传递各种企业运营数据和信息，提供高效获取信息、整合信息的有效途径；③ 打通全企业业务流程过程，优化业务流程，提升业务协作能力，提高运营效率、增强业务的灵活性和透明度；④ 降低服务成本，提升快速适应市场的能力，应对端到端新的服务模式。

企业信息门户就是指在互联网的环境下，把各种应用系统、数据资源和网络资源统一集成起来，根据每个用户使用特点和角色的不同，形成个性化的应用界面，并通过事件和消息的处理及传输把用户有机地联系在一起。它不仅仅局限于建立一个企业网站，提供一些企业、产品和服务信息，更重要的是要求企业能实现多业务系统的集成、能对客户的各种要求做出快速响应，并且能对整个供应链进行统一管理。与面向公众的信息门户相比，企业信息门户肩负着企业最重要的使命——为企业客户的投资增值创建最高效率的业务模式，其功能和特性都围绕着企业间竞争所需的一切高效率而生成，其最突出的特性就是对信息交流的实时双向性的要求。在此基础上，随着具体功能的增加则可区分出不同企业信息门户应用的水平。

企业信息门户可以实现信息集成、信息分类、个性化展示和系统资源管理的集成，实现纵向的企业内部的协同应用和横向的企业间协同应用。在企业信息门户的具体呈现方式上，则从过去 IT 应用系统的“以业务功能为主”的模式转变为“以人为本”的应用模式，通过一个整合的、个性化的场景界面，企业中不同人员可以方便地处理各自的任务、访问相关的数据和信息。

9.4.2 知识门户

企业知识门户（EKP，Enterprise Knowledge Portal）是一种增强的企业信息门户，它把个性化的以浏览器为基础的接口、结构化的数据管理、非结构化的内容管理、协作功能集成为一个服务于知识运作和知识管理的系统^[2, 3]。企业知识门户的特点为：① 目标是进行知识形成、知识集成和知识管理；② 要提供和管理对所提供的信息进行验证的信息；③ 要提供业务信息，以及元信息（这些元信息是指能够依靠它来进行决策的）；④ 能把知识与纯信息区分开；⑤ 能提供从信息产生知识的功能。

企业知识门户的结构是以知识工作者为中心，面向知识工作流的，它以个性化的桌面系统为基础，但还需要知识服务器来支持。企业知识门户是基于信息门户与知识管理的，是目前流行的多项 Web 技术的综合应用，既包括对外的信息门户，也包括对内的知识管理。它应该是单点无缝接入的，其结构如图 9.1 所示。

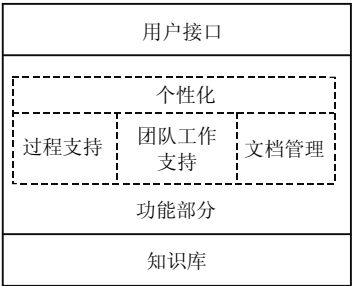


图 9.1 企业知识门户的结构图

从图中可以看出，最前端的是标准化的以浏览器为基础的用户接口，其中可能再挂上一个导引模块。用户接口是知识工作者日常工作所涉及相关主题内容的统一入口。

功能部分一般包括个性化、过程支持、团队工作支持和文档管理四个模块。个性化模块提供用户根据自己工作要求定制所需要的应用与衔接功能，其中包括：个人的信箱，工作日程，查询的记录，人员名录，“推”送信息的名单，个人复制材料等。过程支持模块和团队工作支持模块是企业知识门户最大的特色。在过程支持模块中，工作流、工作清单、项目管理信息等都是重要的内容；在团队工作支持模块中，包含电子邮件、讨论组、聊天室、消息板、电子会议以及会议计划安排等。其中还有一个重要的内容，就是

专家联络表,通过它可以找到某一专门领域的专家,他们具有宝贵的隐性知识。文档管理模块包括接入控制、搜索、导航、内容订阅、文档编辑、文档共享和版本控制等功能,其中最重要的是搜索与版本控制功能。

知识库中保存的知识,则要根据知识工作者所担任的工作而定。例如,从事项目管理的,知识库中保存的是项目文档和以往的经验。从事生产管理的,知识库中保存的则是新闻、报告、供应商和发展趋势等信息。从事客户关系管理的,知识库中保存的则是公司的信息、客户信息,以及与客户接触等情况。从事人力资源管理的,知识库中保存的则是有关人员的教育、经验、技艺等方面的知识以及有关教育培训的知识等。

知识工作者可以通过企业知识门户方便地了解最新消息,当天的工作内容,完成这些工作所需的知识等,可以实时地与工作团队中的其他成员取得联系,寻找到能够提供帮助的专家或者快速链接到相关的知识。企业知识门户的使用对象是知识工作者,它的建立和使用可以大大提高企业范围内的知识共享,充分利用企业现有资源,并由此提高企业知识工作者的工作效率。

这种充分利用企业现有资源的方式是企业提高竞争力的关键。事实上,有效地利用企业已有的知识,提高工作效率和质量,一直都是企业管理所追求的目标。对企业来说,知识的冗余无疑是一种极大的浪费。而企业知识门户就是要挑战时空限制,信息传播缓慢、隐性知识丢失等企业知识利用不充分的状况,跨越横亘在人与知识之间的障碍。作为企业知识管理基础设施中的重要组成部分,企业知识门户提供了知识管理服务平台,并由此在企业经营管理中扮演着重要的角色。当用户在审阅、阅读或引用某一文档时,可以通过企业知识门户发现与上下文相关的人员(也就是能够提供正确信息的专家)和事物(包括内容与信息),这为解决关键业务难题提供了虚拟“场所”,使企业和企业的知识工作者都能够清楚地知道企业所拥有的知识,并在业务中善用所知,提高企业员工的技能素质,进而提高企业的创新能力、响应能力和生产效率。尤其是在企业需要在尽可能短的时间内做出重要决定和采取决定性的行动时,企业知识门户将发挥重要的作用。它可以帮助企业在单一的界面上快速搜索并发现存有知识的地方和掌握知识的人,并加以有效利用,使企业能够更加信心十足地参与市场竞争。

总而言之,企业知识门户是一个能使企业各部门知识工作者之间的知识共享和交流更加流畅的平台。这里的“知识”不仅包括数据库存储的、文档承载的、企业政策和业务过程等显性知识,甚至还包括存在于知识工作者头脑中的工作经验和专业技能等隐性知识。其建立的目的在于:实现企业范围内的知识管理和共享,充分利用企业现有资源,并由此提高企业知识工作者的工作效率,增强企业竞争力。

9.5 知识泵与个性化推荐系统

在知识系统中,知识的收集与分发是一项很重要的工作,知识泵和个性化推荐系统

是两种有效的工具。

9.5.1 知识泵

知识泵这一名称，是从有关组织记忆的研究中引申出来的^[8]。在讨论到组织记忆的类型时，上述文献按照知识收集与知识分发的主动程度，将其分为四类：① 被动收集和被动分发，这被形象地称为“知识阁楼”。在这里，组织记忆指只做档案保存工作。这类组织记忆最容易实现。② 被动收集和主动分发，称为“知识的发布者”。它指把收集到的知识以摘要、新闻报道形式分发给知识工作者。③ 主动收集和被动分发，称为“知识海绵”。组织尽可能使“记忆”完备，但由个人按需要取用。④ 主动收集和主动分发，称为“知识泵”。这是最复杂的一种类型，从理论上来说，它应该包含组织中所有的知识，并能使这些知识发挥作用。

从功能上来说，知识泵是一种引擎，用来沟通组织中的知识流动和应用。它把人员、文档、过程连接起来，并提供正式和非正式的路线图。其所以叫做知识泵，是因为它集知识的收集、分发、检索为一体，而且能够用统计工具对收集的知识进行追踪。通过知识应用的流程进行推断，以改进知识泵的工作，向用户提供社会网络、实践社团、能力和记忆内容的部分路线图。

知识泵在知识系统中的地位可以从图 9.2 中看出。它是沟通知识来源和路线图的工具。图中上端左右都是知识源，左边的是可编码的知识，存在于文档之中；右边是社会网络，包含了专家网络以及与知识有关的社会关系资源，是隐性的，包含意会性知识。该图上段中间部分则是通常的知识管理环节，进行知识存储、评价、注释以及检索、分发和建议等。

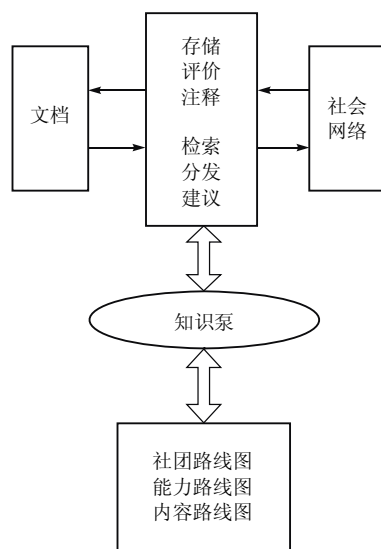


图 9.2 知识泵在知识系统中的地位

在知识系统中, 一项很重要的工作是进行信息的过滤。由于检索得到的信息常常是大量的, 其中有一些是无关的, 应该加以过滤。过滤的方法有两大类, 一类是靠个人的认知能力, 另一类是靠集体的认知。

依靠个人的内容过滤, 涉及文本、图像、声音等的分类方法, 这里最为困难的问题是怎样使自动生成的分类与用户个人的分类方式匹配。再一个困难是怎样对具有不同结构的多种数据库进行查询。

依靠集体过滤的其中一种是所谓协同过滤。协同过滤充分利用现有的社会网络关系, 利用团队的共识作为过滤依据。先由每个人在浏览过一些文档后提出建议, 然后由知识泵集中各人建议, 然后进行汇总。实现它有两种方式, 一种是人工的, 将大家的着重点拿来作标准; 另一种是自动的, 即利用统计算法, 以个人偏好的相关性为基础, 形成建议。

在具体实现知识泵的时候, 要考虑尽可能使用通用的模块, 使其具有较好的可移植性。还要使它有较好的用户易用性, 如检索结果最好就用一般的浏览器显示出来。此外应该使它具有立即得到结果的功能, 使得用户减少操作步骤, 也便于新手较快地掌握使用。

在知识泵系统中, 对于一些常用的文档, 应该可以由用户事先按照需要的迫切程度加以排列, 在调用时可以最先得得到最急需的。此外, 应该有一个记录, 记载过去知识泵都提供了哪些建议, 并且使用户易于找到相关的建议。

综上所述, 知识泵不同于一般查询检索系统的地方, 是它在知识工作过程中, 通过对用户的反复查询、搜索进行的统计和分析, 将各种路线图加以精确化, 使后来的用户从前面已经进行的查询检索过程中得到一些启发和引导, 加快知识获取速度。在信息过滤方面也是突破个人的偏好而吸收团队的共识, 以及其他用户的查询历史经验。因此可以说, 知识泵是一种具有学习功能的系统。它可以不断通过学习改进知识工作。

9.5.2 个性化推荐系统

个性化推荐系统是电子商务中信息过滤的一种有效工具。网络所带来的便捷的信息传递和信息服务推动着电子商务的蓬勃发展, 人们在逐渐享受由此带来巨大惊喜的同时, 也面临着从传统购物方式向网络虚拟购物方式转变的挑战。大型电子商务系统中海量信息的同时呈现, 一方面使用户很难从中发现自己感兴趣的部分, 另一方面也使得大量少人问津的信息成为网络中的“暗信息”, 无法被一般用户获取。个性化推荐系统通过建立用户与商品之间的二元关系, 利用已有的选择过程或相似性关系挖掘每个用户潜在感兴趣的商品对象, 进而进行个性化推荐, 其本质就是信息过滤^[9]。个性化推荐系统不仅在电子商务中具有重要的应用价值, 而且也是一个非常值得研究的科学问题。

个性化推荐系统一般由推荐商品对象、用户和推荐方法构成。推荐系统的功能就是为不同的用户依据用户的历史选择(评价)信息提供不同的推荐商品对象。推荐方法是

推荐系统最核心的部分，这在很大程度上决定了推荐系统性能的优劣。一般情况下，主流的推荐方法有协同过滤推荐、基于内容的推荐和基于网络结构的推荐三种。

(1) 协同过滤推荐的出发点是任何用户的兴趣都不是孤立的，应处于某个群体所关心的兴趣当中，如果某些用户对一些项目的评分比较相似，则他们对其他项目的评分也比较相似。协同过滤推荐方法实现的基本思想是采用某种技术找到目标用户的若干最近邻居（与目标用户有相似兴趣的用户），然后根据最近邻居对目标项目的评分产生推荐，把预测评分值最高的多项商品对象作为目标用户的推荐列表。协同过滤推荐系统可以通过显式评分或隐式评分两种方式取得用户对项目的兴趣程度。显式评分是指推荐系统需要用户直接对某些项目进行评分；而隐式评分则是通过用户的使用日志来获得。例如，如果一个用户在电子商务网站上购买了某本书，则说明该用户喜欢这本书。协同过滤根据与目标用户相似兴趣偏好的其他用户对某商品对象的观点来判断该商品对象对目标用户是否有价值，进而决定是否将其推荐给目标用户，从而缓解信息超载。协同过滤不需要分析商品对象的特征属性，对推荐对象没有特殊要求，能处理非结构化的复杂对象，因而已经成为实现个性化服务中较成功的方法。但是该方法也面临很多问题，如冷启动问题（如何对新用户进行推荐、如何推荐新产品给用户）、评分稀疏性问题以及算法可扩展性问题等。

(2) 基于内容的推荐主要是基于商品对象的内容信息作出推荐的方法。该方法虽然不需要依据用户对项目的评价意见，但更多地需要用机器学习和文本挖掘方法分别对用户和商品建立配置文件。基于内容的用户配置文件需要有用户的历史数据，用户配置文件可能随着用户的偏好改变而发生变化。通过分析已经购买（或浏览）过的商品对象的内容信息，建立或更新用户的配置文件，系统可以比较用户与商品对象配置文件的相似度，并直接向用户推荐与其配置文件最相似的商品对象。基于内容的推荐方法要求推荐商品对象具有较好的结构化信息，其内容要容易获取以便于形成有意义的特征，而且用户的偏好也需要以结构化的形成表达。基于内容的推荐技术在分析商品对象的属性及其相关性时可以脱机进行，因而推荐响应时间快，可以不受冷启动问题和评分稀疏性问题的约束，能推荐新出现的产品和非流行的商品，能够发现隐藏的“暗信息”。缺点是难以区分商品对象信息的品质和风格，并且不可避免地受到信息抽取技术的约束，从而在应用上受到了一定的限制。另外，如果一个系统只推荐与用户的配置文件高度相关的商品对象，那么推荐给用户的商品只能是与他之前购买过的商品非常相似的商品，推荐的多样性很难保证。

(3) 基于网络结构的推荐方法不考虑用户和商品对象的内容特征，而仅仅把它们看成是抽象的节点，其算法利用的所有信息都蕴含在用户和商品对象的选择关系之中。可以把商品对象集合看作 $O=\{O_1, O_2, \dots, O_n\}$ ，而用户集合看作 $U=\{U_1, U_2, \dots, U_m\}$ ，这样就得

到一个二部图 (Bipartite Graph) 关联网络 $N=\langle U, O, E \rangle$, 其中 E 表示该二部图的边集合, 其中的一条边 $e_{ij}=1$ 表示用户 U_i 选择 (评价) 了商品对象 O_j , 否则 $e_{ij}=0$ 。推荐系统的二部图关联网络如图 9.3 所示。二部图关联网络分成用户层和对象层, 通过用户对商品对象的选择 (评价) 关系联系起来。对于任意目标用户 U_i , 推荐算法的目的就是把所有 U_i 没有选择过的商品对象按照 U_i 喜欢的程度排序, 并且把排名靠前的那些商品对象推荐给 U_i 。

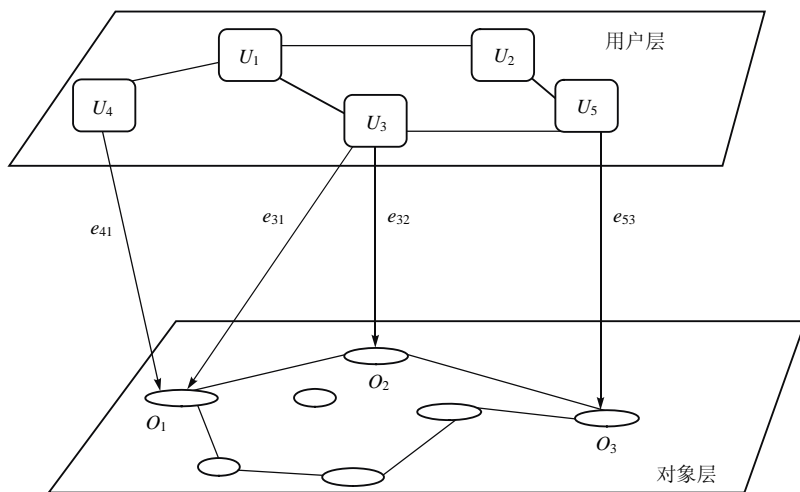


图 9.3 推荐系统的二部图关联网络

基于网络结构的推荐方法开辟了个性化推荐系统研究的新方向。然而, 该方法也面临着冷启动问题。利用网络结构的推荐方法首先需要建立用户—商品对象二部图关联网络。新用户或新产品刚进入系统时, 由于没有任何选择或被选信息, 系统无法与其他用户或商品对象建立关联网络, 因此也无法启动基于网络结构的推荐方法。如果考虑新用户或新商品的配置文件, 可以在一定程度上解决冷启动问题。

9.6 群件与计算机支持协同工作

由于计算机网络与通信技术的发展, 使得许多人可以通过计算机系统在不同的地点协同工作, 交流信息与知识。近年来, 这方面受到了很多人的注意, 在技术上也有很大发展。人们把支持群体工作的计算机硬/软件系统称为群件 (Groupware), 而把这个领域称为计算机支持协同工作 (CSCW, Computer Supported Cooperative Work) ^[2,3]。

9.6.1 群件

群件是支持群体工作的。从广泛的意义上来看, 任何一个有一定规模的系统, 都是在支持群体 (甚至可以认为是在支持整个企业) 工作。但本书说的群体工作, 则是指那

些联系比较密切，经常进行交流沟通的工作。

群件有许多种类型。由于它支持的群体中各人所在的地点不同以及时间上的不同，群件可按下面的时空矩阵来分类（这种划分所用的名称在不同著作中也不同，有的把同一地点与不同地点分别称为本地与远程，把同时与不同时称为同步与异步，但这不妨碍人们理解它们的特点）。

为了便于理解，先在矩阵中填入日常生活中不用现代化信息工具时的相应活动（见表 9.3），后面再看在各种时空条件下怎样使用信息工具。

表 9.3 群体工作的类型

<div>空</div> <div>时</div>	同 一 地 点	不 同 地 点
	面对谈话	电话
同时		
不同时	文件传阅	信件

如果按功能分，群件包括三大类：① 以计算机为媒介的沟通；② 会议和决策支持；③ 分担应用任务或对象。其中第一类只是进行语言或书面沟通。第二类则是为了促成共识；第三类是共同完成一项工作，或者共同使用某一物件，如在建筑工地上用软尺量一个距离，软尺由两个人各执一端，进行测量，结果由一个人口头告诉另一个人，就是日常生活中不用计算机的最简单的协同工作的例子。这种协作可以用图 9.4 来表示，其中 P 是参与的人，A 是共同使用的物件（在计算机系统中可能是一个应用软件），相互之间各种沟通。

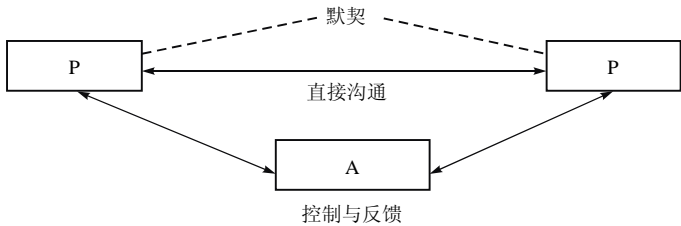


图 9.4 日常生活中不用计算机的最简单的协同工作关系

上面三种功能是当前群件的三种主要用途，也可能在一个系统中兼有两种功能。下面对这三种功能的实现方式作些简要介绍。

1. 以计算机为媒介的沟通

以计算机为媒介的沟通包括电子邮件、电子布告板（BBS）、可视通信和可视会议四种方式。

（1）电子邮件目前已经应用较广，全世界目前已有数以亿计的用户。它与普通邮政

系统相似,也有信件、信封、信箱和邮局等概念,只是全用电子手段来实现罢了。它建立在通信网络上,为每一用户建立一个信箱,给出信箱地址(信箱由一定的存储区和管理软件组成),用户对自己的信箱用钥匙(口令)可以开启。计算机通过调制解调器接入普通电话通信网或直接通过计算机局部网络与电子邮局接通。电子邮件的管理可以是集中式的,由一个称为电子邮局的主机管理信箱与转发邮件;也可以是分布式的,在各用户的计算机上运行。过去的电子邮政是通过文字传信,现在开始使用语音邮件,可以传递发送人的声音,表达人的感情。比起电话来,不会因受话人不在或线路不通而无法对话,又因其不是实时(同步)传送而可大大减少费用。

(2) 电子布告板(BBS)是一种广播式的电子邮件系统,面向更广泛的用户,它可以定期定内容栏目发表信息,如为公众服务的天气预报、交通时刻表、行情和广告等。在一个企业、事业单位内部,可发布公告和通知等。

(3) 可视通信早在十多年前就已经出现了,但由于视频信号传送需要宽频带的特殊线路,在一般通信线路上无法实现,所以一直未能推广。近年来由于综合业务数字网(ISDN)的投入运行,传输率提高了许多倍,加上数据压缩技术的使用,使得可以在一个较小的显示屏上看到通话方的形象,以每秒数帧的速度变化。而在局部网上,可以做到以更大的窗口显示和有更快变化的速度。随着未来宽带综合业务数字网络的开发与投入运行,这方面将会有更大的进步。

可视电话的每一端,除了像普通电话那样有耳机和送话器外,还得有显像器和摄像机。显像器可用一般阴极射线管或平面显示器件(如等离子发光显示、液晶显示、电致发光显示等),后者更适合可视电话一些。可视电话有窄带与宽带之分,前者信号带宽从几千赫到几十千赫,有的只能传送静止图像;后者信号带宽从几百千赫到几兆赫,虽然动态效果好,但费用高。两者都有黑白与彩色两种。

(4) 可视会议可使处于不同地点的人通过计算机与通信系统进行信息交流,它与后面将要讲到的支持决策的会议不同之处在于,后者位于会议室内而这里可用桌上型计算机系统。它可以被看作是可视电话的进一步发展,但它有更多的功能,如还可通过文字档案、图片的交流,进行更深入的讨论,如将X光片传到远方的医生处,进行会诊,将图纸传给远方工程师,讨论方案细节。这种系统一般要使用计算机屏幕,上面有窗口显示对方人物或环境的图像,还可通过软件打开另外的窗口显示图表和文字。因此要求传输的带宽较宽,相当于可视电话的宽带类型。这类系统存在的一个问题是摄像头的视角有限,如果摄录范围大就不清楚,如果移动摄取重点人物,录音又很难兼顾,总之是很难照顾周全。它仅在单人对单人会谈的情况下效果较好。值得提出的是,尽管这种系统效果还不尽如人意,但比起出差到远处又花时间又花钱,费时效率低的情况,系统还是很有优越性的。

2. 会议和决策支持

会议和决策支持可以举出三种典型用法。

(1) 用来作为议论的工具，这是在同一地点而不要求同步参与的情况。参与的每个人既可以把自己的想法、意见和方案放到系统中去让别人看，也可以在系统中看到别人的想法、意见和方案（软件要做到使彼此无妨碍，互不干扰）。然后逐渐取得共识，最后将意见统一到一定的方案上去。这里的不同步，是说不但可在某一时间内讨论，还可留待下次找时间继续讨论。这类系统多用于在同一办公室工作（或者可以在同一办公室讨论问题）的群体。

(2) 用于会议室，这种会议室叫做电子会议室或决策室。室中一端挂有大屏幕显示器，与会者坐在会议桌前，能够清楚看到大屏幕的显示内容。同时，每个人面前还有一个终端。在开会时，与会者可要求发表意见，在屏幕上展示出自己的看法，也可调用某一文件或某一图像，其他人则可以针对显示的内容提出自己的看法，也可打开画面表述自己的意见。画面可以切换，也可以并列显示，这要由专人或会议主持人来加以控制。通过系统，还可进行意见征询，投票表决等。这种形式主要用于同一地点同步工作的。

(3) 共用工作面。这种方式在设计工作中应用极具优越性，它既可以在同一地点到同步工作，也可以在不同地点进行同步工作。每个参与者都可为自己的终端，但终端显示屏上显示相同的内容，也可有共用的大型显示屏，这就像大家在一张图纸上你一笔我一笔共同完成一个设计草稿。如果不在一处办公，还有语音联系功能。当然也可调看其他文件。

3. 分担应用任务式对象

分担应用任务或对象可以举出一些典型应用实例。例如，共用编辑器、共同创建系统、共用一台计算机以及通过某一对象进行沟通等。

在介绍了上述各种形式之后，现在就可能把它们放到时空矩阵中去了。如果回过来再看图 9.4，并把上面三大类型的作用归纳进去，就得到图 9.5。该图中除了把每种联系按上述类型划分之外，还多了一条间接沟通的渠道。

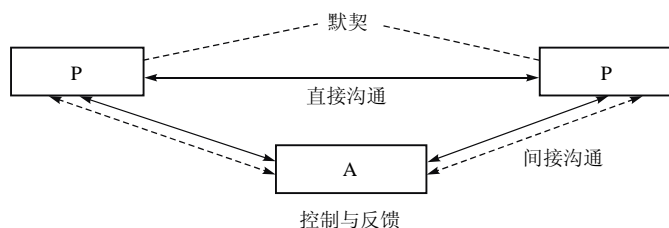


图 9.5 计算机支持协同工作中的关系

9.6.2 计算机支持协同工作

计算机支持协同工作（CSCW）是一个新兴的领域，对于它的定义，尚无众所公认的说法，这里尝试着作如下的定义：计算机支持协同工作，就是利用计算机硬/软件系统与相应的方法和机制，来支持一个群体，协同完成一项有着共同目标的任务^[2,3]。

这里应该说明的是：这种支持不仅依赖于计算机和通信的硬/软件，还依赖于一定的机制与方法。对于一个群体来说，支持的仅是他们工作中具有共同目标的那一部分。因为整个计算机系统对一个企业、一个组织都是支持的，说得更具体点，一个分时多用户集中式系统也可以被认为是服务于一个群体，但其工作并不一定都是紧密联系、具有相同目标的，所以不能算群体工作支持。

表 9.4 计算机支持协同工作

空 时	同 一 地 点	不 同 地 点
同时	电子会议室	可视电话，可视会议
不同时	议论工具	电子邮件，语音邮件

群件的各种类型与工作方式，就是支持协同工作的基础，它们虽然从时间与空间两方面各自具备特点，但从支持共同协作这一点都是一致的。

计算机支持协同工作的产生与发展，是有它的历史必然性的。

人类生活在社会上，经常需要进行沟通，尤其是一个组织之内或者组织之间，经常有需要为完成某一共同目标而协同完成的工作任务。社会越发展，这种活动的广度越广、深度越深。例如，经济全球化，使得许多活动需要跨越地区进行联系、沟通、会商甚至共同做出决定；有些需要当面议定，因此长途旅行是不可避免的；有些则通过信函、电报和电话方式。由于业务的与日俱增，使得交通运输与邮电业务也随之发展，但总觉得跟不上发展需要。中国近年来的情况深刻地说明了这一点。

这种协同工作需要信息沟通，形式也越来越多，过去只限于书面文字或者口头语言，现在则还需要图表、照片甚至视像。过去传送速度过低，人们也不敢奢望传送数量巨大的信息，随着通信工具的发展，人们希望传送的信息量也越来越大。

以上都是从需要方面考虑的。从可能性方面来看，由于通信技术与计算机技术的迅速发展与广泛应用，克服了许多由于空间和时间造成的障碍，使得原来许多不能想像的事现在也能实现了，千里之遥也能像在一室之内面对面地交谈，一方写出的文字、画出的图形，瞬息之间就传到了对方，使人觉得世界变得更小了，人与人之间的距离变得更近了。

计算机技术的发展也正是向这样一个方向发展，从单机单用户到单机多用户，再到多机系统和计算机网络，网络跨度伸展到一个地区、一个国家甚至整个地球。计算机从

互连、互操作到协同工作的发展为这种可能性提供了技术基础。

计算机支持协同工作包括两个组成部分，一个是技术部分，一个是群体工作过程部分（非技术部分）。技术部分包括通信系统、共用工作空间系统、共用信息系统和群体活动支持系统四个系统；群体工作过程部分则包括个人因素、组织因素、群体工作设计和群体动力学研究四个因素。

下面分别对它们之中的一些内容加以研究。

1. 技术部分的四个系统

1) 通信系统

一个群体的成员，平常在一起可以进行面对面的交谈，彼此可以看得见，摸得着，谈话毫无拘束，可以随时插话、打断，并可用表情或手势、形体表示赞成或反对。当他们各处一地时，希望计算机与通信工具能支持他们达到同样功能，这对技术工具提出了极高的要求，一般无法全面做到，但可基本上达到沟通的要求。目前，可用的工具有电子邮件系统、实时桌面视频会议系统和大型公共屏幕系统等。

2) 共用工作空间系统

人们平常在一起工作时，会共用一个文件、一块黑板、一张草图或一张地图进行讨论交流，现在则需使用远程共享屏幕系统、共享大屏幕系统和智能型电子黑板等。

3) 共用信息系统

由于群体中各成员在一起工作时，需要一个共用的信息库，将在活动中所产生的共用信息，也补充到信息库中，所以需要能够支持系统完成信息输入、存储、检索和运算等操作，而信息类型应该是多样的，包括文本、数字、图形、图像和声音等。现在这类可用的系统有多媒体多用户超文本系统、共享光盘系统和多用户分布式数据库系统。

4) 群体活动支持系统

一个群体要想配合得很好，必须在群体工作的目标和工作步骤上取得共识，这是保证相互配合的前提。要想统一目标、协调步骤、严格实施，需要通过工具进行讨论协商，这些工具包括：工作程序与工作流程支持系统、群体工作方法支持系统、群体工作协调系统、合作著述系统、思想生成支持系统、目标确定与排序系统和群决策支持系统。

上面列举的技术因素中，通信部分占有很重要的地位，特别是多媒体通信技术。目前，由于宽带通信网尚处在发展初期，所以还不能普遍使用，一旦在传输带宽和数据压缩两方面有了突破，就会使系统有所改观。

2. 群体工作过程部分四个因素

1) 个人因素

计算机支持协同工作，既有人与计算机交互的一面，又有人与人通过计算机进行沟

通的一面。后一个方面既包含了人-机交互作用（只是这里的“人”应该是复数），更涉及到人与人的关系，不仅有个人心理学因素的影响，还有社会心理学因素的影响。

群体成员的知识、技能和工作风格等，对作为一个组织的整体效能有很大影响，因此需要考虑个人的下列因素：

（1）人类沟通的特征，包括交谈方法，语言表达方式，交往的举动等，对这些有彻底了解将有助于确定一种共同的语言与交流方式。

（2）个人工作模式，包括工作态度、工作方法和工作习惯等。

（3）支持群体的人-机接口形式。这与个人使用时不同，要兼顾群体各个成员，而且在操作时有权力限制。

2) 组织因素

组织因素包括研究一个群体的集体特征：集体知识与智慧的表达方式、组织的设计和组织的组织的管理。

3) 群体工作设计

作为一个群体，对它的整体工作也需要有效地进行规划和组织，包括工作目标的确定，工作规程的设计和业务流程的设计等。

4) 群体动力学研究

群体动力学研究包括：群体协作过程研究、群体行为研究和群体成员行为研究。

这里面研究的是一个组织内部的人怎样行动，以保证群体任务的完成。它要研究一个群体的规模，协作时间跨度，接近程度（面对面还是异地），工作过程如何等，只有把这些研究透彻后，才能确定如何安排技术工具。因此还要研究社会心理学诸因素，如领导与被领导的关系，人际关系等。

这些因素与技术因素比起来，似乎更是无形的，如果说技术因素是硬因素，是看得见的，那么，这些则是软因素，常常被人忽略。而系统应用得成功与否，与它们有很大关系。

例如，群体成员之间或群体之间的沟通模式，可以有点对点、一对多（包括星形、环形和网络形）各种类型。决定选用何种类型的因素，不仅是物理地理位置和通信工具的限制，更重要的是组织形式、工作责权和业务运作规程，当然还包括人们的工作习惯、人际关系等诸多因素，因此必须综合加以考虑。

目前，由于计算机支持协同工作尚在发展初期，人们的注意力多半还放在工具的研究、开发和使用上。随着技术的日益成熟，跟着来的便是上述群体工作过程与协作机制的研究，特别是在知识运作和管理上，这些过程和机制更是决定性的因素。

由于这种协作在空间上有跨度，时间上有跨度，业务活动领域也有跨度，从市场到生产、运输、销售、财务等都有联系。而且它与现有的一些管理方法如全面质量控制等

也有联系,所以常常需要全面考虑,重新设计业务过程,关于这个问题,将在后面讨论。

最后再谈一谈计算机支持协同工作的效益表现在什么方面以及对各类人员的影响。

计算机支持协同工作的效益主要表现在:① 不在一处的群体成员可以不受地域限制直接进行信息交流,节省了时间与旅费;② 不在一处工作的人员可以同时同步协调工作,提高了工作效率;③ 可以提高会议效率,减少会议开支;④ 使决策能得到更多的人参与,提高了决策的质量;⑤ 共享信息与知识,提高了集体能力与智慧;⑥ 秘书工作的某些方面得到改善提高;⑦ 更便于领导检查督促下属的工作;⑧ 提高了文书档案工作水平。

计算机支持协同工作对领导来说可以提高工作效率,减少会议,改善决策质量,更有效地监督整个组织运行;对专业人员说来,能扩大他们的信息收集能力和交流机会;对文秘人员说来,可以改进他们的文书档案工作。总而言之,会提高每个成员的工作质量。

但这里有一个前提,就是不仅在技术上,而且(更重要的是)在组织和业务活动上要充分研究人们日常的工作方式,以及使用计算机后新的工作方式的变化及其对人们的影响,从各个方面调动有关人员的积极性,使他们主动参加设计,这才能使系统真正有用,使用户觉得这是须臾不可离的有效工具。如果不注意这一点,只顾了技术一面,就难免重蹈过去信息系统应用初期开发失败的覆辙。

9.7 Agent 及其在知识管理系统中的应用

Agent 技术为构建分布式知识管理系统提供了一种新的工具。

9.7.1 Agent 技术简介

Agent 这一概念目前还没有公认的确切定义,不同的文献对 Agent 的含义和功能解释各不相同。在信息技术领域,可以将 Agent 看作是有传感器和效应器、处于某一环境中的实体。Agent 通过传感器感知环境,通过效应器作用于环境,能运用自己所拥有的知识进行问题求解,还能与其他 Agent 进行信息交流并协同工作^[10]。

Agent 具有如下基本特性:① 自主性,亦称自治性,即 Agent 能够控制自身的行为,其行为是主动的、自发的、有目标和意图的。② 反应性,即能够感知环境,并借助自身的行为结果对环境做出适当反映。③ 社会性,一个 Agent 一般不能在环境中单独存在,而要与其他 Agent 在同一环境中协同工作。而协作就要协商,要协商就要进行信息交流,信息交流的方式是相互通信。

针对不同的应用领域,研究者设计了多种类型的 Agent,但总体上可分为反应型、

思考型（或认知型）和混合型三种类型。反应型 Agent 的设计者依据行为主义的观点，按照“感知—动作”模型来建立 Agent。反应型 Agent 能够对环境进行感知和监视，并能做出必要的反应。反应型 Agent 虽然对环境的变化有很快的响应速度，但智能程度较低，往往与硬件结合，多用于自动控制和系统仿真等领域。思考型 Agent 是一种基于知识的系统，用符号人工智能的方法来实现 Agent 的表示和推理。思考型 Agent 也被称为 BDI Agent，即有信念（Belief）、愿望（Desire）和意图（Intention）的 Agent。这是目前关于 Agent 的研究中最典型的智能型 Agent，或自治 Agent。思考型 Agent 的典型应用是在万维网上自动收集信息的软件 Agent。混合型 Agent 的结构分为两层，底层是反应层，不采用符号进行表示和推理，可快速响应并处理外部环境的突发性变化，通常具有较高的优先级；高层采用符号人工智能方法进行规划、推理和决策。混合型 Agent 综合了前两者的优点，既有高响应速度，又有较高的智能和灵活性。

从 Agent 的特性可以看出，Agent 的一个显著特点就是它的社会性。所以，Agent 的应用主要是以多个 Agent 协作的形式出现。因而多 Agent 系统（MAS, Multi Agent System）就成为 Agent 技术的一个重点研究课题。另一方面，MAS 又与分布式系统密切相关，所以 MAS 也是分布式人工智能的基本内容之一。多 Agent 系统是一个松散耦合的 Agent 网络，这些 Agent 通过交互、协作进行问题求解。其中的每一个 Agent 都是自主的，它们可以由不同的设计方法和通过不同的语言开发而成，因而可能是完全异质的。多 Agent 系统具有如下特征：① 每个 Agent 拥有解决问题的不完全的信息或能力；② 没有系统全局控制；③ 数据是分散的；④ 计算是异步的。

Agent 技术是一种新的智能技术，关于它的研究方兴未艾。目前的热点课题主要是 Agent 的理论模型、多 Agent 系统及其开发应用。Agent 技术虽然是一种新的智能技术，但它与传统的人工智能技术并不是截然分开的。事实上，Agent 技术与传统的人工智能技术是互相渗透、相辅相成的。一方面，在 Agent 的设计中要用到许多传统的人工智能技术，如模式识别、机器学习、知识表示、机器推理、自然语言理解等；另一方面，有了 Agent 概念以后，传统的人工智能技术又可在 Agent 技术的支持下，提高到一个新的水平。

9.7.2 基于多 Agent 的知识管理系统

Agent 技术在知识管理中的应用源于人们认识到知识管理系统（KMS）与 MAS 之间具有很多的相似性。在 MAS 中，Agent 主要应用于一些动态的环境中，每个 Agent 的行为和推理是由其对环境中实时情况的感知决定的。类似于 MAS，KMS 也可以被看作是一个分布式的系统，其中的每个成员可能具有各自不同的目标，通过协作来完成他们的目标及实现组织的整体目标。在这种环境中，通信和协调的作用是至关重要的。另外，

还要求系统对参与成员及增加共享知识数量具有可扩展性。具体来说,在 KMS 中应用 MAS 主要有以下原因:KMS 中具有资源、问题解决能力和职责的分布性;现存组织结构的完整性和成员的自治性需要得以维持;KMS 中的交互十分复杂,涉及到协商、信息共享和协调等;为使系统设计具有动态性,KMS 要求知识利用和知识源的功能分离;KMS 需求是不断变化的,因此要求系统具有足够的灵活性,能够及时响应外部环境的变化,能够及时主动地把握机会。

MAS 在 KMS 中的应用包含两个方面:一方面,MAS 用于知识管理系统建模运行的组织环境;另一方面,软件 Agent 用于实现知识管理系统的功能。下面根据文献[11]的工作,简要介绍一下第二个方面,即使用 Agent 作为知识管理实现工具的基本建模单元,支持用户的各种活动。

知识管理系统是组织知识管理的实施平台,是一个对知识进行创造、获取、整理、传递、共享,进而创造出新知识的完整的管理系统。KMS 必须同时包含对显性知识和隐性知识进行管理的工具。除此之外,KMS 还应提供对协同工作的支持。根据对 KMS 构建目标的分析,文献[11]提出了一个基于多 Agent 的知识管理系统体系结构,如图 9.6 所示。

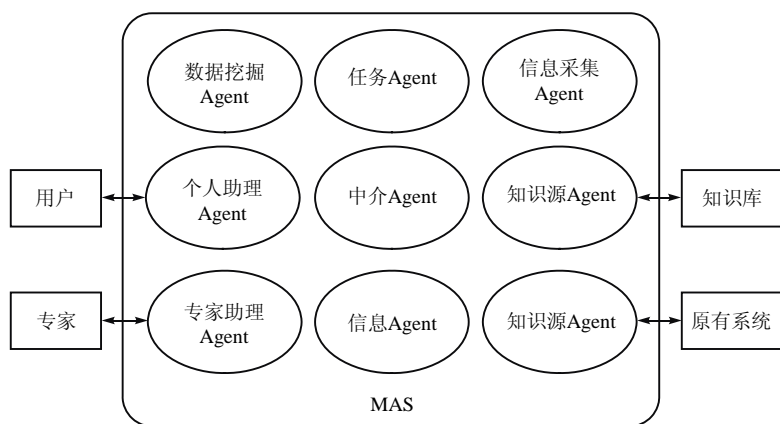


图 9.6 基于多 Agent 的知识管理系统体系结构

基于多 Agent 的知识管理系统体系结构中设计了以下六种类型的 Agent。

(1) 个人助理 Agent/专家助理 Agent。个人助理 Agent 为用户提供与系统交互的接口。它代表用户的偏好和需求,允许用户对知识进行个性化定制。它不仅是被动地完成用户交给它的任务,还能够自动跟踪用户的偏好及需求,适时向用户主动推荐合适的知识源。另外,当用户在进行协同工作时还能自动进行一些文档保存及整理工作,从而减轻了用户的负担。专家助理 Agent 是一类特殊的个人助理 Agent,它除了具有上述功能外,还具有知识导入功能,可以辅助专家将其头脑中的隐性知识提取出来,经过处理后

存储到知识库中。

(2) 信息 Agent。信息 Agent 提供对异构的信息源的智能访问，它具有与信息源相关的模型、资源选择策略和冲突解决机制。它能对外部请求和它们的执行顺序进行推理。信息 Agent 可以从其他 Agent 接收信息。它的主要目标是：回应对信息源的快速查询；应答重复的周期性查询，将结果传送到请求者；监控信息源的变化。

(3) 任务 Agent。任务 Agent 接收用户提交的任务，对其进行分析，确定所需知识，然后与其他 Agent 进行交互以获取所需知识。还可以对任务执行情况及知识利用情况进行监视。具有自学习能力，可以从任务执行过程中获取知识，更新内部状态，维护其任务知识库。

(4) 知识源 Agent。知识源 Agent 负责维护知识源，对其中存储的知识进行描述，并从接受到的请求中提取出相关信息和进行查询。知识源包括企业的各种知识库、数据库及群件系统等。

(5) 数据挖掘 Agent/信息采集 Agent。这两类 Agent 负责系统的知识获取。数据挖掘 Agent 从组织内部挖掘知识。它封装了多种数据挖掘算法，负责从各种已有信息源中进行知识挖掘，并存入相应的知识库中。信息采集 Agent 从企业外部各种知识源（如万维网等）中采集知识，并加以保存。

(6) 中介 Agent。中介 Agent 为其他 Agent 提供各种中介信息服务。它拥有各个 Agent 的注册信息以及组织的知识地图。它为其他 Agent 提供各种可用资源的信息，还可以建议具有共同兴趣或目标的用户进行交流、合作。

多 Agent 的知识管理系统体系结构中的各 Agent 在 KMS 中担任不同的角色，按照各尽所能的原则，协同地完成 KMS 的不同功能，如图 9.7 所示。Agent 之间的协作是以完成一定任务为目标的。

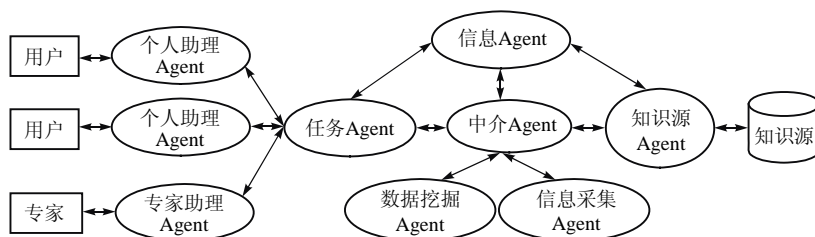


图 9.7 多 Agent 协作

用户通过其个人助理 Agent 向任务 Agent 提交一个任务，任务 Agent 对该任务进行分析，确定完成该任务所需的知识，然后与信息 Agent 进行通信，请求其访问知识源以获取所需知识。信息 Agent 向中介 Agent 查询具有这些所需知识的知识源，中介 Agent 返回的知识源数量可能不止一个。信息 Agent 再与维护这些知识源的知识源 Agent 交互

以获取所需知识,最后将知识返回给任务 Agent 执行任务,并由任务 Agent 对任务的执行情况进行监视,任务完成后经由个人助理 Agent 将结果提交给用户。

用户对所需知识的查询可由相应 Agent 通过交互来完成。若在现有知识源中无法找到用户所需的知识,则系统可以根据知识地图定位到具有该方面知识的专家,并将该专家的位置、专业知识等信息传递给用户的个人助理 Agent,以便建立用户与专家之间的链接,使用户能够与专家直接进行交流,从而实现隐性知识的共享。这些功能主要是由中介 Agent 完成的。除此之外,中介 Agent 还可以建议具有共同兴趣或目标的用户进行交流与合作,最大限度地促进隐性知识在组织中共享。

组织中的知识获取任务主要是由数据挖掘 Agent 和信息采集 Agent 完成的。两类 Agent 接收用户关于知识获取的请求,与用户进行交互,建立个性化的知识获取需求模型,然后从企业内部信息源或外部信息源进行启发式的知识获取。最终,用户将获取到的知识根据组织知识地图存储到适当的知识库中,或直接用来支持决策。

由于企业原有 MIS 等信息系统持续运行所积累的企业各个方面的历史数据是企业知识库的主要知识来源,同时,原有信息系统运行所依赖的计算机系统也为 KMS 的建立提供了必需的物质基础。因此,构建企业 KMS 应以原有信息系统为基础。但是在建立 KMS 时需要保证企业原有信息系统的正常运转。在上述基于多 Agent 的知识管理系统体系结构中,企业原有信息系统与 KMS 的集成主要是通过知识源 Agent 实现的。知识源 Agent 实际上是一个包装 Agent (Wrapper Agent),它通过对外部软件系统进行包装,将其引入到多 Agent 系统中。多 Agent 系统中的其他 Agent 通过与包装 Agent 进行交互来访问该外部系统。MAS 中的其他 Agent 通过 Agent 通信语言将访问命令传给包装 Agent,包装 Agent 对接收的消息进行解析,转换为对外部软件系统的调用命令,再执行调用,然后将结果返回给请求的 Agent。包装 Agent 为 MAS 中其他 Agent 访问外部软件系统提供了一个统一的方法,也为企业原有系统与 KMS 的动态集成提供了一个有效的方法。



思考与讨论题

- (1) 如何利用博客与维基技术进行个人或组织的知识管理?
- (2) 什么是信息门户?什么是知识门户?信息门户与知识门户之间的联系与区别是什么?
- (3) 什么是知识系?什么是个性化推荐?知识系与个性化推荐系统之间有何关系?
- (4) 什么是群件?计算机支持协同工作包括哪些技术因素和非技术因素?



参考文献

- [1] 钟义信. 机器知行学原理: 信息、知识、智能的转换与统一理论[M]. 北京: 科学出版社, 2007.
- [2] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [3] 王众托. 知识管理[M]. 北京: 科学出版社, 2009.
- [4] 谢希仁. 计算机网络[M]. 大连: 大连理工大学出版社, 2001.
- [5] 施卫民, 欧启忠. BLOG 及其知识服务探析[J]. 科技管理研究, 2008, (5): 220-222.
- [6] 王志华, 陈益君. 维基: 基于网络环境下的一种新的信息交流方式[J]. 现代情报, 2007, (5): 53-57.
- [7] 焦玉英, 黄达. 基于 Wiki 的信息服务研究[J]. 情报理论与实践, 2008, 31(2): 226-230.
- [8] Borghoff U, Pareschi R. Information Technology for Knowledge Management[M]. Berlin: Springer-Verlag, 1998.
- [9] 刘建国, 周涛, 汪秉宏. 个性化推荐系统研究进展[J]. 自然科学进展, 2009, 19(1): 1-15.
- [10] 倪志伟, 李锋刚, 毛雪岷. 智能管理技术与方法[M]. 北京: 科学出版社, 2007.
- [11] 蒋翠清, 辛龙潮, 丁勇. 基于多 Agent 的知识管理系统模型研究[J]. 情报杂志, 2007, (2): 56-58.

第 10 章

知识系统的工作过程分析

内容提要

本章介绍知识系统在运作层次上的工作过程，共有 8 个环节，即知识需求与审计、知识采集与编码、知识吸收与利用、知识保存与保护、知识传播与共享、知识转化与生成、知识学习与重用、知识集成与创新。

本章重点

- 知识系统的概念
- 知识系统的工作过程



10.1 知识系统工作过程描述

知识系统中的工作过程可以分为知识运作过程和知识管理过程两类，这两类过程形成了一个两层结构，如图 10.1 所示。知识运作过程是在生产、研究开发或者营销一线的人员为了本身的工作需要而去收集知识或者生成新知识的过程。知识管理过程则是对知识的运作进行管理、提供有利于知识应用和知识创新环境的过程。两者既有区别，又是紧密联系着的。

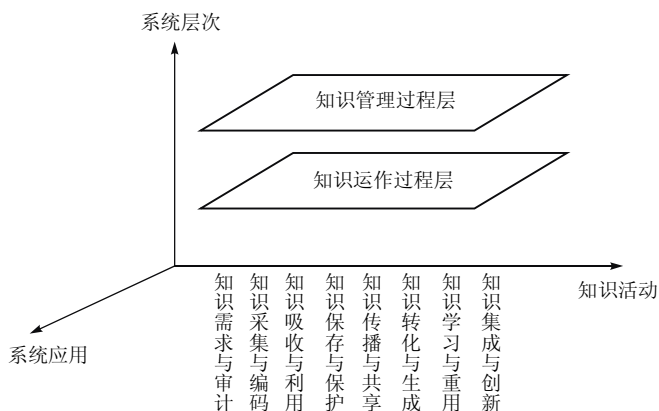


图 10.1 知识系统的工作过程

本章将沿着知识系统的运作过程，来讨论其中的各项活动及其管理方法。这个过程中包含了知识从识别、采集到利用和创新的各種知识活动，我们将沿着下面这样一个顺序来讨论：知识需求与审计、知识采集与编码、知识吸收与利用、知识保存与保护、知识传播与共享、知识转化与生成、知识学习与重用、知识集成与创新过程。这样依次进行研究，实际上也就对知识系统的主要功能进行了讨论。

这里需要注意的是，在分阶段、分功能研究知识系统的运作过程时，要时时刻刻记住它是一个整体，每一阶段、每种功能都是为整体目标服务、受整体制约的。

10.2 知识需求与审计

知识需求分析和知识审计是知识系统运作过程的起点。只有先明确了企业的知识需求和知识差距，才能使知识系统启动起来。

10.2.1 知识需求

所谓知识需求是指为了完成某项工作或活动所需要的知识。明确了知识需求有助于回答这类问题：企业需要什么样的知识？这些知识可以从哪里获得？

按照不同的分类标准，知识需求可以分为个体需求和群体需求、目标需求和一般需

求、职业需求和专业需求、正式知识需求和非正式知识需求、核心知识、先进知识和革新知识需求, 内部知识需求和外部知识需求六种需求。

1. 个体需求和群体需求

知识需求的主体是人, 人的知识需求有个性与共性之分。个体需求千差万别, 最具有个性化特征, 满足个体需求是知识提供的一贯要求。群体需求指政府机构、企业、事业单位等群体的知识需求, 群体对知识的需求具有整体性的差异, 即各群体有不同的知识需求。

2. 目标需求和一般需求

用户知识需求总是带有一定的目的性, 一般需求表现为增长知识、科学研究、教学参考、消遣娱乐等, 在满足一般需求的同时, 必须兼顾特殊的目标需求。比如, 可以向用户提供他们所需的决策、方案、创意、技巧、程序和软件等知识, 也可提供系统性的参考资料。

3. 职业需求和专业需求

因职业和专业的不同, 则有职业需求与专业需求之分。有多少职业就有多少职业需求, 如行政人员、军人、医生、教师和学生等的知识需求就有很大差别。专业需求则是对某一专业领域的知识需求, 一般来说, 专业需求与用户所从事的职业紧密相关。

4. 正式知识需求和非正式知识需求

正式知识是那些被证明为正确、可靠的理论化知识体系, 多收录于权威性工具书(如词典、手册、百科全书等)和经典性教科书中; 非正式知识包括研究报告、学术论文以及有争论性的观点等, 其研究成果还需进一步论证。正式知识需求一般指为掌握知识而进行知识储备的需求, 非正式知识需求指为利用研究现状, 以及进行知识创新的需求。

5. 核心知识、先进知识和革新知识需求

核心知识是企业生存所需的最小范围和最低限度的知识; 先进知识构成了企业在行业内的竞争优势; 革新知识是指使企业远远领先于其所在行业的其他竞争对手, 从而使企业与其竞争对手严格区分开来的知识。

6. 内部知识需求和外部知识需求

企业内部知识需求包括两个方面: ① 企业在自身发展过程中, 进行扩张、发展、优化、结构变革等过程中产生的知识需求, 如企业产品线的扩展、企业要素的优化、战略目标的调整、管理水平提高、信息系统建设等; ② 企业运营中出现问题时产生的知识需求, 包括职责混乱、效率低下、士气低落、人员流失和市场份额下滑等问题出现时的需求。

企业外部知识需求也包括两个方面: ① 企业需要把握外部环境中存在某些机会时产

生的知识需求。这种机会包括新技术的兴起、新市场的出现、用户新需求、新政策公布、竞争对手决策失误等；② 企业需要应对外部环境中出现的某些威胁时产生的知识需求。这些威胁包括对企业不利的政策出台、供货渠道或者销售渠道的变动、产品市场的变迁、环境中某些突发性变化等。

对企业知识需求的分析，可以采取以下步骤：

（1）识别和预测知识需求。识别和预测企业在目前或未来将要开展的项目、事务过程中需要哪些知识，明确企业目前或未来需要的知识。其中，确定所需知识的类型是识别和预测知识需求的关键。

（2）分析企业目前的知识状态。通过知识调查和知识清点，可以收集到关于企业知识的质量、数量、存取、使用、共享和有效性等方面的信息，确定企业实际的和潜在的知识财富，包括对显性知识的统计、标引和分类，可以客观地识别和评估企业的知识。企业目前知识状态分析的结果可以用知识地图表示，它可清楚地揭示企业知识资源的类型、特征、知识之间的相互关系。

（3）分析知识缺口，并确定相应的弥补策略。在以上两步分析的基础上，分析企业知识缺口。企业需要的知识与企业目前所掌握的知识的差异称为知识缺口，这可分为内部和外部知识缺口两种。由企业自设的内部目标和自有资源之间的差异构成的缺口称为内部知识缺口，企业内部与外部竞争者在知识上的差异构成了外部知识缺口。针对不同的知识缺口，企业应采用不同的知识缺口弥补策略。

那么需要的知识从哪里获得呢？可以考虑从企业的内部知识源和外部知识源中获取。内部知识源包括企业内部 R&D 部门及营销、生产等其他部门；外部知识源包括供应商、分销商、研究机构、竞争对手、大学和交易会，等等。外部知识源类型具体来说可分为市场类知识源、机构类知识源、标准类知识源和其他类知识源四大类十六小类，如表 10.1 所示。

表 10.1 外部知识资源类型^[4]

知识资源类型	
市场类知识源	1. 用户和分销商
	2. 设备、原材料等供应商或承包商
	3. 竞争对手
	4. 咨询顾问
	5. 独立的商业实验室和研发企业
机构类知识源	6. 大学或其他高等教育机构
	7. 政府的研究组织（如研究所等）
	8. 其他公共部门（如生产力促进中心）
	9. 私人研究机构

续表

知识资源类型	
标准类知识源	10. 技术标准
	11. 健康、安全标准和规定
	12. 环境标准和规定
其他类知识源	13. 专业论坛、学术会议
	14. 行业协会
	15. 技术、贸易出版物和计算机数据库
	16. 交易会、博览会

10.2.2 知识审计

知识审计是企业对其自身拥有的知识资源进行系统、科学地考察和评估的一种有效手段。换句话说，知识审计可识别企业核心的信息和知识需求，分析知识差距和知识流等，并判断它们对经营目标的影响。知识审计有助于回答这类问题：为了解决特定的问题，企业已经有了哪些知识，缺乏哪些知识，谁需要这些知识，他们如何使用这些知识。

1. 知识审计对象及范围

知识审计的对象包括知识主体（人）、知识客体（即知识本身）和知识环境三个方面。

（1）知识主体。人是知识的拥有者，也是知识的接受者。在这部分，一方面要审计人的知识流程、彼此的关系等内容；另一方面还要熟悉企业各部门的工作内容与目标，以便确定知识需求。需要提及的一点是，这里所说的人不仅指组织内的知识工作者，还包括了客户、供应商等企业价值链上的知识主体。

（2）知识客体有显性知识和隐性知识之分。显性知识存在于数据库、文档等载体之中，而隐性知识存在于人的头脑中。所以，审计知识时，不仅要审计结构知识（即人走后，留在公司的知识，基本上相当于显性知识），也要审计隐性知识。

（3）知识环境。知识环境对知识活动起着推动或阻碍的作用。知识环境有两个方面：技术环境与企业环境。技术环境主要是指信息技术，也可以理解为 IT 设施。企业环境包括企业设计和企业文化，企业设计指企业的流程和结构，企业文化间接地反映了知识共享的文化和企业的价值观。

知识审计的范围包含两个层次：一是已有的、企业所知道的知识，即知识基础；二是企业需要明确知道自己所需要的，企业所缺乏的知识，即知识需求。只有知道拥有哪些知识，才有可能找到知识存储与传播的最有效的方法；也只有知道知识需求，才能找出知识差距，知识管理项目才能收到预期的效果。

2. 知识审计方法

知识审计是一个新兴的研究领域，还没有成熟的审计方法。目前，比较流行的有

AnnHylton 提出的 HyAKAudit 方法、Delphi 集团的知识审计方法——KM2 方法、Liebowitz 等学者提出的审计方法等。上述几种知识审计方法，都采用了问卷调查和面对面访谈的方式，通过分析调查结果形成知识审计报告。此外，可以使用已开发出来的评估工具，不过这些工具不一定完全适合每一个企业，需要做适应性的改进和修正。

无论运用上述什么方法，执行的关键步骤基本上是一致的。一般而言，问卷调查和访谈相结合是常采用的方法。因为知识的特性使得知识价值的评估不仅仅根据投入与产出，还在于知识主体的洞察力。企业中的每个人对知识都有自己的观点和角度，什么样的知识是有价值的，在什么状况下有价值。也就是说，独立的个体决定怎样、什么时候在他的工作中使用隐性知识，而知识的产生、存在和特征依赖于特定的人——拥有隐性知识的人，企业中大部分的知识是以隐性的方式存在。所以知识审计要以人为中心，对知识的审计需要从人对知识的理解入手。此外，知识审计的目的是最大程度地发现、开发、利用潜藏在企业中的隐性知识，并尽量把这些知识编码化。只有进行深入的调查，才能揭示隐性知识的分布，了解知识工作者的知识需求。需要注意的是，在设计调查问卷前，需要清楚企业有哪些专业知识领域。

3. 知识审计流程

企业开展知识审计需要依据一定的流程，这个流程一般包含三个阶段。

(1) 准备阶段。确定知识审计的目标，进而确定知识审计的对象、审计的范围、审计的指标和审计报告的结构。只有明确了知识审计的目标，知识审计的过程才有意义，最终完成的知识审计报告才有价值。这个阶段的主要工作是：了解企业的基本情况、了解企业的内部控制体系、初步评价企业的知识资产（大类）等。

(2) 实施阶段。完成准备阶段的工作后，要根据企业的具体情况，选择知识审计的方法。使用的方法决定了企业是否能够准确地诊断知识的状况。知识审计的实施包括数据的收集和分析。

(3) 完成阶段。最后对知识审计得到的数据进行整理和评价，形成审计建议和意见，编写知识审计报告，以统一的框架将企业拥有的知识资产文档化。除此之外，一般还要绘制知识地图，描述企业的知识流，找到企业的核心知识，找出企业的知识差距。

10.3 知识采集与编码

为了得到所需要的知识，在确定了的知识领域和已建立的知识储备框架的基础上，需要着手采集知识，并对其进行编码。

10.3.1 知识采集

知识采集是企业知识流动的初始性工作。首先要确立知识采集的目标，然后确定知

识采集的内容和范围,有针对性地进行采集,减少不必要的周折,提高采集知识的效率。

常用的知识采集方式在第 4 章中已经介绍过。

下面研究一下知识的维度。知识中有的已经通过编码呈现在纸面(书籍、文档)上,或者在计算机存储系统中;有的需要从有知识的人那里采集并编码;也有的是隐性知识,只能在使用知识时请具备知识的人来,身临其境地运用他的知识或者与其他人的知识融合而生成集体知识。所以在采集知识时,还要考察一下,应该从哪几个方面来看待知识,这就是知识维度问题。

一般说来,在知识采集时需重视的维度有知识形式化的程度、知识领域、理论与实际的特点和知识领域的成熟度四个方面。

1. 知识形式化的程度

从知识的形式化程度来分,知识可以分为显性知识和隐性知识两类。显性的知识能够以文本、图像、表格以及专家系统的形式表述,可以阅读、解释、讨论和应用,并可以用纸面形式或电子形式保存在知识储备库中,并可通过推理等机制,产生新的有价值的知识。

显性知识在企业内部的来源有:手册、说明书、报表、备忘录、调查报告、管理信息系统和流程图等;外部来源则有:书籍、杂志、财务报告、新闻、研究报告和产业分析等。互联网、学术会议、趋势分析、咨询报告和典型介绍等也都是来源。

隐性知识存在于个人、组织的记忆中,在研究、讨论、调查的当时,人们能够进行陈述与交流,但如果不及时把它转化为有形的知识,随后也就丢失了。

2. 知识领域

知识领域可以按照企业系统的内部和外部来区分。

企业系统外部有下列诸方面的知识:市场、竞争、顾客、产品、服务和供应等;企业系统内部则有下列诸方面的知识:过程、测度(财务、生产、顾客和雇员)、管理、专长、技术、结构(组织、团队和工作角色)、激励(发展、授权和报酬)和文化等。

3. 理论与实际的特点

在理论与实际方面,知识可分为理论知识与实际知识两类。理论知识包括基本概念、原理、模型和学说,都是由该领域中的专家通过多年的经验与实践抽象和概括出来的。实际知识则包括应用原理,惯用规则,经验以及其他一些诀窍,都是应用于日常工作的。有时专家在遇到问题时,会用一些走捷径的方法来解决,这种方法是人们从经验中得出来的,常常是不经意地掌握的,很难利用知识形成的方法变成有意识的知识。

4. 知识领域的成熟度

知识领域的成熟度是与知识的结构化程度相关联的，一般分成良结构化的、半结构化的和非结构化的三类。

(1) 良结构化的，如算法、公式、理论、框架体系和分析过程等，这些都是可以建模、推理的。

(2) 半结构化的，如判断、启发式思路和决策规则等，这些可用基于规则的方法来推理。

(3) 非结构化的，缺少甚至没有理论，如新的领域和案例形成的经验等，可用基于事例的推理。

有形的知识收集起来以后，还要对其进行筛选，检查所获得的知识是否已经存在于企业的知识储备库之中，新的知识值不值得放到企业的知识储备库中去？

一般说来，知识的选择要考虑下面几个方面的因素：

- (1) 知识在该知识领域中的针对性与价值；
- (2) 知识的准确性；
- (3) 发现重复的知识并加以清除；
- (4) 寻找缺少的知识；
- (5) 提高不确定知识的可信度；
- (6) 解决不同知识带有的矛盾；
- (7) 对无法消解矛盾的知识建立多种框架和视角。

从单纯的技术观点来看，加入知识储备的知识应该是正确而且一致的。这从逻辑上来看是合理的，但事实上并不必要。有矛盾时可以请专家来讨论矛盾是否能消解，实在消解不了时，可以建立多种视角和知识框架结构。有时候，特别是在一个复杂的知识领域中，许多问题尚无众所公认的定论的时候，保留不同的认识有时反而有利于发现新的线索。

10.3.2 知识编码

知识编码，是在知识采集结束后，对其进行识别、分类、提取、组织和标准化等一系列加工和处理的过程，使无序的知识转化为有序的知识，使其便于公开、共享和交流，并能够通过信息手段进行传递。知识编码是提高知识显性化程度，将知识从无序到有序、隐性到显性的过程。

知识编码一般包括电子化、结构化和标准化的内容。

1. 电子化

企业知识的编码可以从书面资料的电子化工作入手，将企业内部和外部的相关知识

通过文字输入或扫描的方式形成电子文件，建立一定的子目录系统，进行计算机管理。文件电子化、数字化是知识编码的基础，有利于知识的存取、积累、交流与共享，建立企业数据库，进而为建立企业内部网络 and 知识仓库打下基础。

2. 结构化

将电子化、数字化的知识按照企业知识的分类进行整合，分别形成不同的人员、组织和数据的知识仓库，三者之间通过知识的流动形成程序性知识。其中人员知识库包括人员的求职、招聘、离职及其他管理的知识；组织的知识库包括客户知识、政府知识、媒介知识等，数据的知识库包括档案、资料 and 文件等；程序性知识库包括项目的开发、流程和管理等。

3. 标准化

由于编码的目的是为了实现标准化和规范化，因此，知识编码一定要建立统一的编码库。只有建立统一的编码库，制订编码的制度和使用规范，才能有效地整合企业的知识，使工作流程规范统一，减少重复性工作、提高效率和准确度。

知识编码的困难在于隐性知识难以编码。因为显性知识易于通过计算机进行整理和存储，而隐性知识存储于员工的大脑里，是员工的个人经验。在进行知识编码时，一方面可以建立知识仓库，以收集各种经验、备选的技术方案以及各种用于支持决策的知识。通过模式识别、优化算法和人工智能等方法，对成千上万的信息、知识加以分类，以知识条目的形式串联保存在记忆单元中，提供决策支持。同时，建立承载隐性知识的案例库。当企业或者员工再次面临相同或相似环境的时候，能通过阅读、分析以前的案例，从中得到需要的隐性知识来解决面临的问题。另一方面，建立一个包含相互联系的任务、文档和事件等元素所构成的知识地图，这个地图允许用户在浏览时发表评论，对地图进行更新，不断地改变地图中各个元素之间的联系，使知识地图逐步趋向完善。知识地图的另一种使用方法是描述企业流程中的知识，将业务流程中的知识流通过图表的方式展现出来，整合关键知识，以达到提高生产效率的目的。

10.4 知识吸收与利用

当人们得到知识后，能否吸收而变成自己或企业的知识，要受到多方面的制约。

首先是要看新的知识能否与已有的知识相适合。这有两种极端情况，一种是使新知识适应原有的知识，另一种是使原有知识适应新知识。但大多数情况是介乎两者之间，也就是相互改变来适应。

其次是要看新知识能否与团队、企业的观念和习惯等相一致。

再就是要看新知识能不能够落到实处，产生实际效果。从吸收了知识到真正能使用

知识，期间还有许多前提条件和准备工作，所以还得做一个估计。

怎样评估吸收的程度，可以从两方面去看。一方面是看知识嵌入个人或企业的工作过程的程度，另一方面是看能否重复使用。

从知识运作和知识管理的全过程来看，每一个环节都应该考虑到，怎样使知识得到有效地利用。常常有这样的情况，企业引进了新的技术，对员工进行了新知识的培训，但是新知识却并未得到应用。这样不但工作没有得到改进和提高，反而浪费了资源。

现在没有任何一个企业不使用知识，但有的企业使用了新的知识，就比那些未使用新知识的企业更容易成功。所以，不断获取新的知识并加以利用，是企业今后生存与发展的重要手段之一。知识的利用不是一劳永逸的，当某一类知识从识别、获取到使用的过程完成之后，又会发现需要新的知识，又需要开始再识别、获取到利用的过程，因此知识工作过程是一个不间断的过程。

企业为了使知识发挥效用，应该鼓励人员使用知识。这首先要使企业领导认识到知识的重要性，有利用知识提高自己的核心竞争力的要求；另一方面也要说服教育企业中的人员，克服固步自封的思想与情绪。同时采取种种措施，帮助他们学习新的知识。

造成吸收和使用新知识的障碍，来自不同方面。首先是企业本身的盲目性与短视，总觉得自己已有的知识足以使工作顺利进行，不需要外来的新知识。特别是在有些工作流程已经程式化了的情况下，更不愿意改变。例如，当初德国的汽车制造业者就不愿吸收日本的生产方法。

另一种阻力来自文化方面。常常有一种习惯势力使得人们习惯于他所熟悉的原有的方法和思路，拒绝接受新知识。在这种气氛中，如果有人想要吸收或利用新知识，就会处于一种孤立与受歧视的状态之下。还有的人是觉得吸收别人的知识有损于他本人的尊严而拒绝新知识。还有的是部门领导过于自信，对邻近单位的经验和知识觉得不值一顾。凡此种种，都会使得企业成员在是否吸收新知识的问题上无所措手足。

从领导方面来看，有时候还会由于新知识不是本单位创造的而拒绝吸收和利用。这实际上还是面子问题在作祟。其实只要新知识真正对工作有用，真正能创造效益，不论来自何方都应该一视同仁，当然有时候出自成本的考虑也是合理的。

从向使用者提供知识的侧面来看，应该尽可能提供方便，把知识的使用者看作是客户。使用者总是希望通过最方便的方式得到知识，特别是在急于使用的时候。如果通过一个电话、一次浏览或者一次交谈就能获得所需知识，他当然乐于使用。所以为了促进知识的广为利用，知识管理者应该尽量做到使用户觉得友好，这表现在：

- (1) 简单性，使用户不需要什么准备就能得到并使用。
- (2) 及时性，最好随要随到（JIT, Just in time）。
- (3) 兼容性，可以很容易地转向其他知识源。

(4) 新颖性, 一般说来, 总是知识越新, 用处越大或者越广。

对于知识储备库来说, 应该有对用户友好的人-机界面, 使人们乐于使用。

此外在工作环境布置等方面, 也要设法使工作人员感到取用知识的方便。

最后应该提到的是从知识管理的角度来看, 使用知识的过程决不是孤立地使用了所获得的那部分知识。一般来说, 总要与使用者已有的知识、特别是隐性知识相结合才能解决问题, 因此总有一些自己独创的东西。应该鼓励使用者在干中学, 好好总结, 反过来再充实企业的知识库。

10.5 知识保存与保护

知识只有通过有效的保存和保护, 才能得以传播和共享。个人知识可存储于大脑中, 而企业知识则需存在组织记忆中。

10.5.1 知识存储

下面先解释一下前面提到过的概念——组织记忆。

对于一个人来说, 为了使用和创造知识, 首先需要记忆一些知识。对于一个组织来说, 是否也需要记忆一些知识呢? 显然是需要的。也许有人认为, 组织是由个人组成的, 只要个人记住某些知识, 也就等于组织记住这些知识了, 事实并不是这样, 因为个人记忆的知识, 不是一下子就能被获取、分享、理解和使用的, 个人的知识即使对他本人来说, 也有忘记的时候。对一个组织来说, 如果这个人离开了, 这部分知识也就消失了。所以还得设法建立组织记忆, 尽可能使知识形式化。

问题是组织需要记忆哪些东西? 忘记或忽视哪些东西?

组织记忆的内容应包括下列知识与信息: 知识源的索引与检索技术、计划与调度程序、工作步骤、原则与导则、标准与政策、因果关系模型、信息与数据存储、决策规则、绩效指标及其他数据、业务系统设计、利益攸关者与客户简况(需求、价值观、期望、感受)、产品与服务(特色、功能、价格、销售、修理)、领域人力简况(知识、经验、偏好、兴趣)、领域的卓越成果和现状评估等。

组织记忆除了大量存在的形式化知识(如文档和书刊等)外, 还包括下列五个组成部分:

- (1) 个人的记忆与能力;
- (2) 文化(信念、价值观、符号和情节);
- (3) 变革的过程与系统;
- (4) 生态(环境与改造);

(5) 外部关系(客户、竞争对手、行业群体和政府)。

下面来研究一下组织记忆是如何生成的。本章前面是把知识按领域划分的,因此可以将组织记忆看作是许多领域分科组成的一个网络,它的一个组成部分是存储装置,包括人脑、纸面文档和计算机;另一个组成部分是组织和个人业已形成的习惯、观念。例如,个人就可把他的知识放在纸面或电子文件中,以及放在他的记忆、信念、价值观和心智模型中。

知识的记忆和保持,与知识的表示方法有着密切的关系。在进行知识存储之前,有两个知识维度需要考虑:一是知识的结构,二是推理机制。对用户来说,知识结构是已经确定的;推理是主动的,能对结构进行处理以产生有用的输出,如推论和答案等。

每一种知识结构都有专门的推理机制与它相配合。例如,超文本链接与关键词搜索就能给数字化了的文本增加价值。关系型数据库作为一种知识形式,如果没有相应的查询语言,就不是很有用。又如规则库如果没有推理机的使用,也不能解决任何问题。

关于知识的存储,还需要引入知识储备库(knowledge repository)的概念。知识储备库是一种在线的存放某一专业领域中的知识、经验和文档的计算机设备。知识经过收集、格式化并转换为数字形式后,可以放进知识储备库。它包含如下多种知识结构。

- (1) 字典,包括知识领域中名词的定义、概念和词汇。
- (2) 图像库,包括数字化的图像和视频文件。
- (3) 文本库,包括书籍、期刊、手册和说明书等。
- (4) 文档库,包括超文本。
- (5) 数据库,包括各种类型的数据库。
- (6) 实例库,包括决策与解决问题的典型事例。
- (7) 规则库,包括定义性的知识,决策和解决问题的规则。
- (8) 脚本库,包括事件、过程和典型的行为。
- (9) 对象库,包括概念、实体和对象。
- (10) 过程库,包括工作流程。
- (11) 模型库,包括因果关系模型。

以上囊括了几乎所有的知识结构模式,对一个具体的知识储备库来说,只能选择其中若干种。下面着重讨论其中的几个。

(1) 实例库。实例库至少应该包含两个部分,一个是工作实例库,一个是决策准则库。实例库包含将工作经验与教训、成绩与错误都记录下来的文档。它们应该覆盖各个方面。使用实例应该经过应用训练,使之能很快推广到新的情况中去。

(2) 规则库。规则库主要用于决策支持。建库时先要确定影响企业的都有哪些主要

的决策。对决策者来说, 外界的、他人的知识可以用来改进决策质量。一般是先选定决策变量, 然后收集有关的信息。可以从最好的专家那里提取知识和规则, 然后用实例去检验这些规则。

(3) 过程库。这是描述企业如何进行工作的流程图的在线集合。可以把流程分解成一项一项的任务, 但不能分得太细。过程库在进行仿真与预测时是很有用的。应该收集一些成功的工作经验, 加入到知识储备库中。

10.5.2 知识产权保护

对于一个企业而言, 知识产权包含了两层含义, 一是有了知识产权意味着企业有了参与竞争的权利和资格, 有了立足于市场的资本。二是有了知识产权意味着企业无形资产的增加, 进而可以获取效益。基于此在技术还未成型的研究开发阶段, 不仅要强调技术创新的速度, 紧跟市场的脚步, 而且必须树立企业员工的保密意识, 限制接触核心技术的人数, 加强对每一批实验数据和结论的管理控制, 规定保密人员的责任, 必要时签订保密合同。

特别是对于企业的核心知识, 为了保持企业的竞争优势, 需要采取有效的手段在一定的期限内加以保护, 以防止其他企业或个人效仿。

正式的保护方法有专利、商标和版权等。非正式的保护方法有保密、锁定策略、领先市场和复杂设计等。

10.6 知识传播与共享

知识共享是生成组织知识的必要条件, 是企业创新的重要前提。要想进行共享, 首先要使知识能够传播。目前, 企业中的知识传播与共享的程度还很低, 通过人的交谈与纸面上的交流进行传播与共享是最普遍的形式, 但效率较低; 通过信息网络, 利用电子媒体进行传播与共享是最快而最容易进行的方式, 但投资较大。从发展的眼光看来, 后者一定是越来越普遍。

10.6.1 知识传播

知识传播涉及三个方面的问题, 一是传播的技术问题, 二是传播的方式问题, 三是如何查询检索的问题。

1. 知识传播的技术手段

过去知识传播靠的是纸面文件、电话和传真。现在则逐步使用信息技术工具。知识传播的技术手段有如下几种:

(1) 电子邮件是当前使用最为方便的传播工具，它可以是一对一的传播，也可以是一对多的传播。

(2) 企业中的内联网，是当前知识传播最有前途的工具，它一方面可以与互联网相连，很容易把从互联网获得的知识传递给企业内部需要使用知识的人，另一方面也为企业内部的部门与部门之间、人与人之间的知识传递提供了方便。

(3) 群件应该说是最有利于团队进行知识传播的工具，因为它提供了符合团队工作范围和工作习惯的传播模式。

(4) 再进一步就是电子视频会议，特别是远程会议系统能使身处异地的成员像在同一地点那样接受知识的传授。

(5) 智能主体能够为知识传播提供自动化服务。

(6) 组织记忆为上述各种方式提供了大量的知识源。

2. 知识传播的方式

知识传播的方式有推和拉两种。所谓“推”的方式，就是根据知识接收者的偏好和知识需求等信息，向知识源预订有关的内容，知识源会定期或不定期地将信息或知识推送给订户。所谓“拉”的方式，就是知识使用者先去检索，然后按需要将有用的信息或知识拉过来。

3. 知识的查询检索

知识的查询检索在第7章中已有介绍。

为了进行知识检索，应该尽可能将它数字化并存放于计算机系统中去。由于系统中有像关键词检索等工具，因此要比检索纸面上的知识容易。

至于知识的分发，有人建议可给组织记忆加上一种自动分发功能。智能主体（智能代理，也就是软件机器人）可以用来在互联网上搜索有关的知识，提供给感兴趣的人。

在知识的传播方面，还有一个矛盾的问题，就是传播和保密的矛盾。从企业的立场来看，一方面希望某些知识能够广泛传播，以提高员工的水平和能力；另一方面，又有一些企业所独有的知识不希望泄露出去，特别不希望泄露到竞争对手那里，否则自己的竞争优势就会丧失。此外，对知识保密的范围加以恰当界定也是十分必要的。如果保密范围过于宽泛，会使员工的进步受到限制，同时也会使员工感到企业对自己缺乏信任从而降低了对企业的忠诚度和信任度。

10.6.2 知识共享

知识共享实际上是一个沟通过程，是指知识所有者与他人分享自己的知识，是知识从个体拥有向群体拥有的转变过程。知识共享的目的是知识创造，实现企业内部的知识增值。

而在企业中进行知识共享,可提升企业的知识价值和知识创造能力,企业的员工或内外部团队、企业内部或跨组织之间,彼此通过各种渠道进行知识交换、交流和讨论,可扩大知识的利用价值并产生知识的效应。

依据正规化程度和企业员工互动方式的不同,知识共享可以分为两种基本模式:正式知识共享和非正式知识共享。正式知识共享的基本特征是,知识共享主要依赖于正式的制度,有着较为固定的内容、程序或形式的约定。因此,企业员工的正式交往,制度化的技术文档整理与知识传播,将知识贡献于企业数据库等成为正式知识共享的主要途径。在这种模式下,企业的正式系统和横向协调成为知识共享的主要载体,知识被有目的地获取并通过正式机制传播。依赖于员工个体社会关系而非制度获取知识,则是非正式知识共享的基本特征。它独立于组织结构、政策和正式协作之外,体现了个体间基于私人关系的一种知识援助,具有典型的非正式特征。相应地,企业各部门及员工之间的非正式交往、兴趣小组、网上论坛、实践社区等则成为非正式知识共享的主要途径。知识在这个过程中以共享价值观和非正式网络为载体被有意或无意地传播。

知识共享还分个体间知识共享和群体间知识共享。个体间知识共享可依赖于 Blog 和 RSS 技术实现,而群体间知识共享则可基于 Wiki 技术实现。相关技术已在第 9 章中介绍过。

知识交流是知识共享的基础。实现企业知识的交流需要有相应的组织结构和文化氛围,要在企业内部提倡知识共享和交流的价值观,企业必须调整价值取向,营造开放而又信任的合作环境,最大限度地提供公共信息资源,增加工作流程的透明度,形成有效的业绩评价与激励机制,为员工实现知识的交流提供方便的渠道,让员工有更多的机会与其他人进行交流。而且这种交流不仅要在具有相同知识结构的人员之间进行,更重要的是要与具有不同知识结构的人员交流,以便在交流中了解对方的经验知识,解决实际问题。同时,在应用知识进行项目开发时,能够直接接收到具有不同知识结构的人员的帮助,以弥补自己知识的不足或知识结构上的欠缺。如果企业内不同组织、部门和个人能够相互交流知识,利用各自的经验,将会给企业带来巨大的知识财富。

在企业中,组建“知识社区”是实现知识交流的有效途径。“知识社区”是指由一些具有共同兴趣、来自相同或不同团体的人们,由于互动的需求所形成的群体。企业中的“知识社区”有如下两种形式。

(1) 实体知识社区。它是由企业以本身核心竞争力的定位来设计和组建的。根据企业发展的目标,将具有各种知识的专业技术人员组织起来,进行定期或不定期的活动,如调查研究、交流会、讲座和开展知识评价等,使参与者从不同的知识结构和知识领域内获得灵感和启迪。

(2) 虚拟知识社区,则是利用网络系统在企业的知识管理平台上开展活动。由于它

没有明确的主题限制，因此往往可以收集到一些企业领导者没有想到的知识，而且也适于全体员工的参与，所涉及的知识范围更广泛。

10.7 知识转化与生成

人们在利用知识的过程中，有时需要将外面获得的信息或从企业内部获得的经验、事例、情节转化为形式化的知识，有时为了创新需要，还需在已有的知识与经验的基础上生成新的知识，因此需要研究知识的转化和生成过程。

10.7.1 知识转化模型

关于组织知识的转化和生成，日本的野中郁次郎（Nonaka I.）和竹内弘高（Takeuchi H.）提出的知识转化和生成模型很有启发意义^[5]。它把知识看作是个人的信念经过检验向真理发展的过程。

这一模型包括三个主要组成部分：一是知识通过社会化、外化、组合和内化的转化过程；二是知识生成的平台，他们把它称为“Ba”，三是知识资产。图 10.2 表示了三者之间的关系。下面分别对这三者加以介绍。

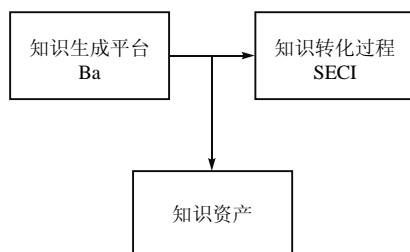


图 10.2 知识转化和生成模型

第一个组成部分阐明了显性知识与隐性知识的转化和丰富充实，个人知识与组织知识的相互转化与生成的过程。

在本书的第 1 章曾经讲到，知识的两大类型：显性知识与隐性知识之间有着许多差别。显性知识能够以形式化语言表述，能够以数据、公式、规范和规则等形式进行共享，容易处理、传递和保存。可以说显性知识是涉及过去的对象和事件，是“彼时彼地”的知识。

隐性知识则是高度因人而异的、很难形式化的知识。它深深地植根于人的行动、工作步骤、习惯、思路、价值观和情绪之中。可以说隐性知识是“此时此地”的知识，是在一个特定的时空背景下的知识，它很难与别人进行交流。

隐性知识有两种，一种是技术性的，包括记忆和诀窍等。能工巧匠可以加工出精品，但他很难将他的经验总结成科学理论。另一类是认知性的，是一种“心智模型”，像一种

观点、信念、图式和视角。它反映了个人对世界的看法和对未来的愿望。

在知识转化与生成模型中有四种知识转化过程，即：

- (1) 由隐性知识到隐性知识的转化过程；
- (2) 由隐性知识到显性知识的转化过程；
- (3) 由显性知识到显性知识的转化过程；
- (4) 由显性知识到隐性知识的转化过程。

通过螺旋式的转化，个人知识逐步聚集并转化为组织知识，又扩大和深化了个人知识，推动了知识的产生和创新活动。

隐性知识到隐性知识的转化称为“社会化过程”。这常常是通过观察、模仿、实际操作和形体上的沟通而不是通过用语言文字的教学来完成的。过去师傅带徒弟传授技艺就是这种方式。要做到隐性知识的传递与共享，人们必须在一起活动。

要做到个人所掌握的隐性知识为他人所共享，必须使个人的思维模式和某些技巧能为别人所接受。有时非正式的聚会、交谈或者在共同完成某项工作的过程中，心领神会地做到了知识的传递。有时可以用讲故事的方式进行这种沟通与传递。这种沟通不仅发生在同事之间，还会发生在产品开发者与顾客之间。

隐性知识到显性知识的转化称为“外化过程”，但这里所称的外化过程与哲学以及心理学中的含义是不同的，这里指的是把有关的不可表达的知识表达出来。隐性知识转化为显性的概念，这个概念可以是一种隐喻，一种象征性的口号或提法，或者是一个模型。这表现在产品或服务创新过程中提出的形象性目标，或者在完成一项新任务时提出的口号。外化过程在知识产生和创新活动中是非常重要的，因为要通过它产生新的概念。

显性知识到显性知识的转化被称为“组合过程”。这可以通过文件、会议、电话交谈和信息网络等媒体，用分类、合并和增补等方法，将分散的概念组合成知识系统。例如，公司把营销中产生的数据加以综合分析，会发现什么产品畅销，什么产品滞销，从而制定出新的销售策略，就是一种组合。

从显性知识到隐性知识的转化被称为“内化过程”。这里的内化，是指将显性知识增补到个人的隐性知识中去。如果显性知识已经载入文件、手册或者某些讲话中去，内化过程就会进行得顺利一些。有时候，人们读了某一位企业家的自传，介绍他的创业经历，从而得到启发和感悟，改变了他的思维方式，也是一种内化。

组织知识就是在上述两类知识的四种交互作用过程中产生的，这常常需要多次循环，形成螺旋式的上升，如图 10.3 所示。

文献[5]中讲到的一个例子很能说明问题。松下电器公司产品开发部在开发家用自动烤面包机时，遇到焙烤过程中的外焦内生问题无法解决，开发人员甚至使用 X 光机对机器揉制与手工揉制的生面团进行透视比较，也没有什么有意义的发现。后来软件开发负

责人田中郁子与几位工程师去大阪国际饭店当学徒，从面包师那里观察体验到面包师在抻开生面团时还要加以拧转，他们把这一隐性知识学到后，经过讨论，形成了概念，在机械结构和操作程序中以设计形式实现，把隐性知识转化为显性知识，实现了外化过程。然后将各种技术知识组合起来，形成样机构型。

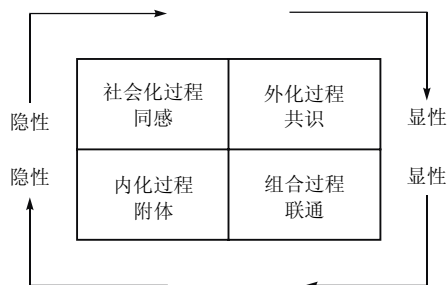


图 10.3 野中的知识转化模型（SECI 模型）

我们是否可以这样说：在创新过程中，由于社会化过程使参与者在隐性知识方面形成默契，通过外化过程，按照创新目的形成概念，达成共识，然后通过组合过程使知识更加丰富与系统化，再通过内化过程使参与者的知识有所增加和深化，能力有所提高，去从事创新成果的实现。这样一个过程是从个人知识到组织知识再到个人知识的螺旋式上升过程。从系统理论的角度看来，这个系统是个自组织系统，特别是共识的达成与新概念的产生，符合自组织的一些原理，可以从自组织系统角度对它进行深入分析。

在为企业进行信息系统的规划时，就有这样的螺旋式上升过程。开始时，作为用户方的企业有关领导和作为开发方的学校方面的信息系统与信息管理专业人员通过接触，介绍情况和初步考察，双方产生同感，都觉得企业的经营管理需要信息化。但这仅仅是从直觉和经验得出的感觉，这时通过交谈，互相交换了对工作的感受，共享了隐性知识，形成了默契，认为这项工作是可以合作进行的。

经过多次相互介绍业务内容，用户方面把原来对信息系统笼统的感觉外化成具体的需求，开发方把对企业中业务的笼统感觉外化成具体的业务流程。双方开始在开发的范围、深度上初步达成共识。

在这个基础上，双方通过调查、讨论、相互介绍，把各自掌握的业务知识与信息技术知识汇总起来（这中间避免不了争论和反复），形成规划知识的组合，逐步形成规划方案。

在双方讨论、修改、重施方案过程中，每个人又从方案的显性知识中得到启发和感受，然后充实到自己的隐性知识中去。这样完成了第一个循环。

在论证、审批规划的过程中，又可进行后面的循环。

在 SECI 模型和理论中，强调的是两种知识的转化和互补，但是在显性知识与隐性知识之间，除了相互转化的过程外，按本书作者的看法，还存在相互融合（fusion）的过

程,融合将会起到更深层次的作用,融合是中国传统哲学的一个特点。我们认为,显性知识在人脑中或物质载体中是成块存在的。人们学到一门知识,如某一门数学,或者数字电路原理,其中包含许多概念,它们按照一定的逻辑关系结合成一个整体。有一部分知识块,由于相互有关而形成一个网络。而隐性知识则是弥漫在一个人的头脑(指个人的)或一群人的头脑之中的。更多的显性知识块之间、显性知识中的某些概念集合之间,是靠隐性知识把它们融合起来的,这是从静态方面来看。

从动态方面来看,隐性知识到显性知识的转化在宏观层面上是一种突变,是一种涌现过程,或者说是一个顿悟过程。但是我们认为,在实际的创新过程中,还有许多转化过程,如图 10.4 所示。既有像图中 A 部分那样,从还不够明晰的知识或经验,通过像专家系统、数据挖掘或文本挖掘这类知识发现方法,变得更加明晰并形成规律,易于言传;也有像图中 B 部分那样,通过讨论从不太能够完全说得出来变得能够说出许多;还有像图中 C 部分那样,从完全缺乏经验到通过实践和观摩变得能够形成一些经验积累。一句话,都是在不同程度上提高知识的显性度和可言传性。

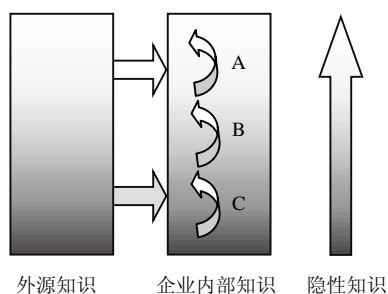


图 10.4 基于知识谱系的知识转化

像图 10.4 中右边那样,直接从完全隐性转化为完全显性的知识只存在于某一些知识点上。如果仅仅通过二分法(如 SECI 模型)来看所有的知识转化,那么就无法解释上面说到的知识的显性度渐变式的提升,如数据挖掘、文本挖掘这一大类知识发现过程与方法了。其实这是一种渐增式的外化过程。当然同样也存在着渐变式的内化过程(图中未画出)。

在整个转化的循环过程中,人和组织是起主导作用的,但是技术工具能够起到一些支持和帮助的作用。下面来研究在各个转化阶段,人与技术手段是怎样协同工作,各自发挥自己的优势的。

在社会化过程中,由于是隐性知识的传递与共享,所以应该有充分的隐性知识资源。除了组织内部有关人员的选择要考虑这一点之外,还可以邀请有丰富经验与间接的外界人士一同来通过正式的和非正式的会议或交谈来沟通隐性知识。也可组织参观考察,开阔眼界,得到第一手直接的感性知识。此外,还可以通过各种渠道从客户或有关的组织

机构搜集一些议论、动态等，从中得到启发。

在这个阶段，一些信息技术工具无法起到重要作用，因为沟通的是隐性知识，需要面对面通过形体或者感受来沟通。但是一些通过形象演示有助于启发和引导的工具，如某些情节与环境的多媒体视频演播，会收到语言文字所不能达到的效果。另外由于远程视频会议系统的应用，能使分散在各处的人们也能面对面地交流。当然也不可忽视其他虽然不是面对面、但仍然能够沟通的工具，如音频会议的作用。

在这个阶段，还可以发现志同道合的人。比如，在为企业开发信息系统时，在初步接触中通过交谈，能够发现企业的领导和专业人员中，有一些具备较好的信息素养，有很强烈的愿望和主动性的未来合作者。发现这种情趣相投的人，也是一种扩大隐性知识资源的办法。

在外化过程中，要进行隐性知识到显性知识的转化。这个过程在知识生成中是最为关键的，因为它要形成新的、明显的概念。这需要通过富有创意的讨论和对话过程。

在这个过程中，除了人与人的当面接触外，一些远程协作的信息工具也能起到一定的作用，它不仅可以使得人际沟通突破空间限制，而且还提供远程观察现场和实物的条件。此外，像网上专题讨论组与论坛等也能起到一定的支持作用。由于概念的形成常常通过一边讨论一边写写画画，因此可以借助于某些信息显示工具作为讨论的园地，或通过把概念的结构关系和层次用图形表示出来的软件工具，起到形象化的描述和梳理的作用。这里还有一个使用工具的好处，就是可以把讨论过程与结果全都记录在案，以备日后查询。

在组合过程中，各种分散的显性知识要加以组织，形成更加完整、复杂的显性知识。这种组合通常是使用文件、图形、数据表格等形式，通过会议、交谈等协同工作方式完成的。也可以借助于远程通信工具来完成。这是由于有一些知识是通过外化过程获得的，而另外一些知识则是通过企业的知识储备库中或者从互联网上获取的，这时一些查询检索工具就显得十分重要。由于现在绝大多数文档已经电子化了，图纸也被计算机辅助设计的图形文件所代替，所以很多组合工作可以借助于信息工具来完成。

在组合过程中，除了要将各类有关知识加以组织成为整体之外，有时候还需要再把知识分解成若干部分，由各有关部门去应用。

一般说来，组合过程分为三步：第一步先将现有的和进一步搜索得到的知识组织起来；第二步是把结果分发给有关人员征求意见；第三步是根据意见加以修改提高，使其具有可操作性。

表面看来，这一过程只是在显性知识的范围内进行加工，但实际上还是会涉及隐性知识的。我们认为，这种组合是一个从局部形成总体的过程，这个新的知识总体应该涌现出新的整体性能。组合要靠人的专业知识，也要靠经验和直觉，甚至灵感。有时候这个组合过程会带来新的意想不到的结果。现在人们认为技术创新有两大类型，一类是原

创性的创新,另一类是系统集成性的创新。在组合过程中就会出现后一种创新。

内化过程是把显性知识依附到个人身上成为隐性知识。这实际上就是“干中学”的过程。知识的内化会开阔、延伸和改善原有的隐性知识。通过共享心智模型,可充实组织的隐性知识。

在这个过程中,内化的知识有的是与具体的实践活动有关联的,将在今后的类似活动中得到应用;另外有一些则是涉及工作方法和策略的,会在今后更广泛的领域内起作用。

在内化过程中,虽然信息技术工具不能直接参与,但是许多辅助学习的工具在这个转化过程中会起到一些潜移默化的作用。例如,人们在观看了一个仿真演示之后,所得到的绝不仅仅是其中提供的显性知识,一些感受、体验都会变成隐性知识。

这四个过程进行多次循环,螺旋上升,就会生成新的组织知识。这种过程会在各个层次中(团队、整个组织和组织之间)进行。个人的隐性知识乃是组织知识生成的基础,但是经过组织活动,加以扩展并凝聚成更高层面上的知识。

10.7.2 知识生成环境

知识的生成需要一个适当的环境,在上述模型中,知识生成平台——Ba 就是这种环境。Ba 的原意是一个场所,是野中等人从一位日本哲学家 Nishida 那里引用过来的,他们把它看作知识共享、知识生成和知识利用的一个背景环境。Ba 可以是一个物理空间(如办公室和分散在各处的工作场所),也可以是一个虚拟空间(如电子邮件和远程会议等),或者精神空间(如共享的经验和思路等)。在这个平台上,最重要的活动是互动。因为知识生成的力量不仅在个人,而且在人与人之间、人与环境之间的互动。平台正是这样的—一个空间。个人能够共享和再生成知识,就因为他处在这样的空间之中。

这样看来,把这样的平台看作是动态的知识要比把它看作是一个含有知识的物理空间更为恰当。而知识生成过程也可以看作是知识生成平台的生成过程。

日本学者称这一平台为 Ba 是有他们的历史渊源的。但这一简称对其他人来说不容易理解,不如称其为知识生成虚拟平台(下面简称平台)。中国学者在钱学森的指引下,提出处理开放复杂巨系统的综合集成方法,使用的综合集成研讨厅这样一种方法和工具,就是这种平台。

知识生成虚拟平台也可以被认为是将知识作为一种资源的活动架构。因为知识本身是无形的,无边界的、动态的,而知识的使用需要把知识资源集中在一定的时间内和一定的空间中。例如,当知识生成之后,具有知识的人员和知识库是组织在一定的空间之内的,这个空间就是知识生成平台 Ba。

知识生成虚拟平台和创新团队有很多相似的地方,都是集中了知识工作者和知识资源来生成新知识。但是它们之间还是有区别的。知识生成虚拟平台的形成和消失完全是

看需要, 参加的成员并不是都很固定; 而创新团队在一定的历史阶段人员是固定的。知识生成虚拟平台的边界是动态的, 而创新团队的边界按照使命与任务是固定的。

知识生成虚拟平台有初始平台、对话平台、系统化平台和作业平台四种类型。

(1) 初始平台, 是对个人来说共享感觉、情绪、经验和心智模型的场所。在这里先是共享隐性知识, 开始知识的生成过程。因为这里人与人之间是面对面的互动, 是全面获得人的形体和心理情绪感觉的唯一途径, 而这正是传递隐性知识的重要因素。

在这里, 人与人之间消除了一些隔阂, 建立起关怀、信任、忠诚和爱心的桥梁。这些都是构成人际知识转化的基础。

(2) 对话平台。通过对话平台, 个人的心智模型和技巧可以转化为一些术语和概念, 也就是隐性知识转化为显性知识。在这一环境中, 对话和再思考是关键活动, 因为通过它们可以共享别人的心智模型, 同时分析自己的心智模型。在对话平台, 人们能超越自己的局限, 并能理性地运用所获得的知识来处理自己的知识。因此在这一平台上个人能把理性与直觉综合起来从而获益。

对话平台的构成比起初始平台有更鲜明的针对性, 因为人们是按照专门知识和能力来选择是否参加这个平台的。

(3) 系统化平台。通过系统化平台要把各种显性知识组合而生成新的系统化的显性知识, 因此一些信息技术工具如文档处理、协同工作环境等会起很大作用, 这一点本章前面已经提到过。因此这个平台的技术环境是主要构成要素。

(4) 作业平台。在这一平台上, 要进行内化过程, 把显性知识转化为隐性知识。由于人们是在实践活动中不断地学习和深化自己的知识, 而不是通过培训教学, 因此通过自己的行动和思考就可连续地把知识内化到个人身上。在这里, 是通过行动来转化知识, 而不像在对话平台是通过思考的过程。这里的活动是即时当地的。

这四个平台与前面提到的四个知识过程是对应的。应该说明的是, 这些平台的概念以及它们之间的联系都是动态的, 有很大的灵活性。

从企业领导或知识总管的角度来看知识的转化与创新, 需要把上述模型中的三个组成部分, 在统一的知识愿景和知识战略的指导下, 充分运转起来。这中间的相互关系可用图 10.5 来表示。

从图 10.5 中可以看出, 根据组织的愿景确定的知识愿景, 是整个过程的原始出发点。企业领导或者知识总管先要对企业的知识资产心中有数。知识总管对知识资产的评估与管理, 应该像财务总监 (CFO) 对企业的有形资产的评估与管理一样, 但是不能把它看作是静止的, 而应从动态的过程着眼。

由于知识是没有边界的, 所以对它的界定还不能仅限于技术、产品和市场等方面。特别是组织的例行工作程序, 组织文化, 都是无形资产。处理的好了会是一种动力, 处

理不好会变成一种阻力。

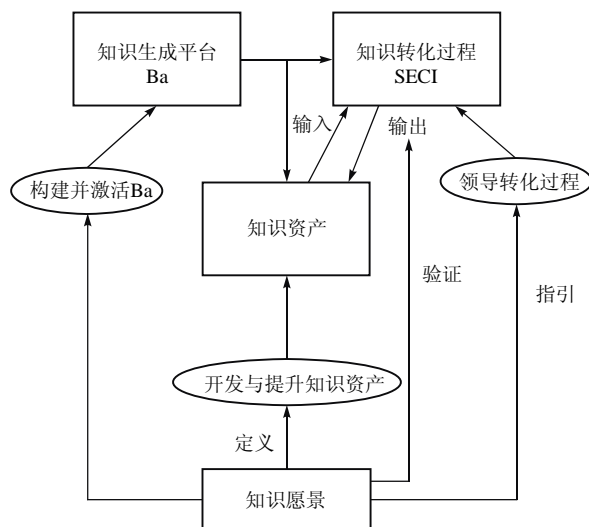


图 10.5 知识生成、转化与资产之间的关系

知识生成虚拟平台的构建，有时候是自动形成的，有时候是自觉形成的。作为领导，为了促进平台建设，应该提供物理空间（如会议室）和信息空间（如计算机网络），促进人员之间的互动，形成骨干队伍。但也还可以按照需要随时临时组织人员进行意见交流。

为了激活平台，发挥作用，除了物质条件外，需要提供一些精神上的条件。

自主性是首先要加以重视的。由于知识的生成不能够按照一种刻板的模式进行，所以应该给组织中的成员以更大的自主权，这样才能增加各成员之间交换知识和利用知识的机会。利用自组织原理而形成的一些团队是发扬自主性的好形式。团队之间的交叉影响也拓宽了知识范围。有些企业使用了应急项目团队的方式，来对外界的变化做出反应。

在组织中形成一种议论纷纷、各抒己见的氛围，对于知识的生成是很有帮助的。这可以让人们不断提出问题，重新认识原来的知识，舍弃以前的见解而重新学习，使得知识转化和生成过程发生急剧变化，让一些新思想和新方案脱颖而出。

野中等人所提出的这一知识转化与生成的整套模型和环境，对于组织知识的生成做出了很好的分析，在当前知识管理领域有很大的影响。但是它也存在一些不足之处。例如，有人批评他们把显性知识与隐性知识更多地作为物件而分得太清，正如本章前面说到的，两种知识之间并不存在鲜明清晰的界限，而应该看成是一个连续的谱。

从野中对知识转化与生成的研究历程来看，他是从提出四个转化过程发展到提出知识转化平台 Ba 的概念。后来又与组织学习紧密联系起来。他多次声明，西方对知识管理的研究过于侧重显性知识，忽视隐性知识的作用，忽视组织知识的作用和生成。他认为自己的思想体现了东方的思维模式，对人与我、主观与客观、身与心的关系做到了统

一。但在强调组织知识作用的同时,对于如何发挥个人创造潜力,如何创造适合个性发展氛围等方面的内容,还缺少深入的考虑。这也许是日本民族文化所决定的。

中国的传统与现代哲学思想中,有许多光辉的成就涉及到知识的转化与生成,如《实践论》中的许多观点对知识生成与转化有很大的指导意义(值得一提的是野中在日本大学的研究生课程参考书中就有《实践论》与《中国革命的战略问题》)。我们还需要充分利用这些宝贵的资源来研究知识创新问题。

10.8 知识学习与重用

知识的重组和新知识的产生与学习是分不开的。个人的知识产生是在一定的需求或启发下,树立一定的目标,通过学习、调查、实验和实践,输入基础知识、专业知识以及相关的信息和知识,通过思维加工,构成一定的知识结构,再经过实践和创造性思维,进行知识重组,从而产生新知识的过程,同时也是学习的过程,其中包括单回路学习过程和双回路学习过程。而组织知识的产生和创新则离不开组织学习。下面重点介绍组织学习的类型。

10.8.1 组织学习

组织学习一般是指建立在组织过去知识和经验记忆的基础上,通过共享信息、知识和各种学习模式而建立的一种学习过程和学习机制。自20世纪80年代提出组织学习的概念到20世纪90年代,国内外学者对组织学习已从心理、政治、经济、社会和文化等多方面进行了研究,把组织学习看作是形成未来企业核心竞争力的重要因素。这种学习机制在于企业能否通过学习提高对产品、技术和管理的创新能力,形成新的特殊资源,再通过对这种资源的有效利用形成企业的核心竞争能力。建立在组织学习机制基础上的管理创新机制是企业形成管理优势的源泉。在组织学习中,每个成员对学习过程和结果都产生着重大的影响,但组织学习绝不是个体学习的简单相加。组织成员与组织之间的交互行为、组织与外部环境相互作用、组织文化的构建都是组织学习的重要特征。

组织学习的实质,不是把散沙式的个人学习装到一个大沙盘里,而是以团队式组织作为学习的主体。而那种个人学习的集合,是一种非凝聚状态的“组织学习”,承袭的依然是提升个人能力就可以对组织发展做出贡献的老观念。他们没有看到个人知识的提升并不能确保组织能力的提升。要实现组织整体能力的提升,必须使经验、知识和技能在不同的部门和人员之间有效传递,而且交汇融合,共同分享。这样才能真正提高组织能力,形成共同愿景。

组织学习的类型包括:

(1) 单循环学习 (single-loop learning)。单循环学习是指发现并改正组织错误, 使得组织能够保持当前政策, 去实现既定目标。单循环学习可以被视作为一种将组织所需的知识、技能或规则植入组织, 同时不改变组织现有形态特征的组织活动。单循环学习亦被称为“低层次学习”、“适应性学习”、“复制式学习”或“非战略性学习”。

(2) 双循环学习 (double-loop learning)。双循环学习是指除了发现并改正组织错误外, 能对组织现有的规范、流程、政策以及目标提出异议和修正。双循环学习包括对组织的学习基础、特殊能力、例行常规进行变革。双循环学习亦被称为“高层次学习”、“创造性学习”、“拓展型学习”或“战略性学习”。战略性学习就是组织与所处环境相适的一种努力过程, 在此基础之上, 组织才能够充分占有各种资源, 并以这些资源去追逐、拓展和实现组织目标。

(3) 第二次学习 (deutero-learning)。当组织知道如何执行单循环学习和双循环学习的时候, 组织就开始了第二次学习的过程。如果组织对于学习无意识的话, 就不会有前两种学习。“只有认识到无知, 才会产生学习的动机”。这意味着在促进学习之前, 需要找到学习导向、方式、流程以及结构等学习促进因素。Nevis 等人找到了七种学习方式和十种影响学习的促进因素。例如, 其中一项促进因素就是, 找出目标业绩与真实业绩之间的差距。认识到这一差距后, 组织就能够想到学习, 以及想到创造适合学习的环境和流程。这同样也会使组织意识到这样一个事实, 冗长的积极反馈或者是沟通交流事实上会阻碍组织学习。

双循环学习与第二次学习都非常关心组织为什么要变革和如何变革, 而单循环学习则被动接受变革, 不做深层思考。Dodgson 指出, 组织学习的类型还取决于组织学习所存在的不同组织部门。不同的学习活动可能出现于不同的组织功能部门, 如研发、设计、工程、制造、市场、行政及销售, 等等。

10.8.2 知识重用

知识系统中的知识重用过程有两层含义: 将相同的领域知识重用到不同的问题求解方法和将相同的问题求解方法重用到不同的应用领域。要想实现系统的知识重用, 仅仅将不同层次的知识区分开是不够的, 还需要一种有效的机制来实现各层次间的灵活配置, 将相互独立的层次紧密地联系在一起, 共同组成一个完整的系统。本体就是实现这一机制的核心。

知识系统中存在不同种类的本体, 这些不同种类的本体的可重用程度是不一样的。知识系统中常用的本体类型有^[6]:

(1) 知识表示本体 (knowledge representation ontology), 指获取用于基于框架语言的表示原语, 这些原语包括类、子类、属性、属性值、关系和公理。

(2) 通用本体 (general ontology), 包含与事物、事件、时间、空间、因果关系、行为和功能等相关的词汇。

(3) 领域本体 (domain ontology), 指给定与一个领域的概念、概念之间的关系、原理与基本法则有关的词汇。它在一个给定的领域是可重用的。

(4) 任务本体 (task ontology), 指提供用于解决某一特定任务的术语, 这个任务可以是领域无关的, 也可以是领域有关的。

(5) 领域-任务本体 (domain-task ontology), 指只能在一个给定领域重用的任务本体。

(6) 方法本体 (method ontology), 指为用于完成某一特定任务的具体推理过程所用到的概念及概念之间的关系提供定义。方法本体和领域本体通常用来定义 PSM (problem-solving methods)。

(7) 应用本体 (application ontology), 包含一个特定应用系统建模的必要知识。

(8) 元本体 (meta-ontology), 也被称为类本体 (generic ontology) 或者核心本体 (core ontology)。它定义了 “part-of” 关系和其上的属性, 这个关系可以用来说明一个装置上的各种组件的集合, 每个组件又可分解成子组件。元本体可以在不同的领域中被重用。

在应用本体开发知识系统时, 本体的可重用性和可用性是一个需要考虑的问题。一般来说, 一个本体的可重用性越好, 那么它的可用性也就越差, 反之亦然。例如, 当使用本体为领域知识建模时, 所用到的本体包括: 知识表示本体、通用本体、元本体 (类本体)、领域本体及应用本体等, 这些本体的可重用程度和可用程度如图 10.6 所示。

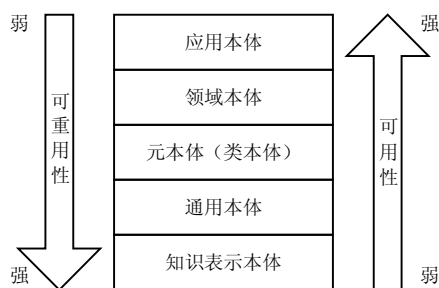


图 10.6 本体的可重用性和可用性

当开发一个新的知识系统时, 如果能充分利用可重用的组件, 就可以更好、更快地完成该系统的开发。知识系统的本体和 PSM 都可作为可重用的组件, 本体关注的是静态的领域知识, 而 PSM 关注的是动态的推理知识。因此, 为了使不同领域和任务的知识及推理行为能被共享和重用, 各种不同的本体和 PSM 被迅速地开发出来。

在开发一个新的知识系统时, 要想重用已经存在的 PSM 和领域知识, 那么必须解决好 PSM 和领域知识的连接问题。给定一个与具体领域无关的方法和一个与任务或方法无

关的领域知识，有三种方法可以将它们连接成一个知识系统：

(1) 重新修改 PSM，使其适合领域知识。

(2) 重新修改领域知识，使其适合 PSM。

(3) 引入一个新的中介，该中介可消除 PSM 和领域知识的差异。把这个中介称为影射关系。

在实际的应用中，方法(1)和方法(2)的实现是比较困难的，只有方法(3)切实可行。在知识系统中引入连结 PSM 和领域知识的影射关系后，系统将有三个可重用的部分：与 PSM 无关的领域知识，与领域知识无关的 PSM，连接 PSM 和领域知识的影射关系集合。

只要实例化影射关系就可以实现影射关系的重用，提高开发的效率。为了避免重用所引起的难题，连接 PSM 与领域知识的影射关系应尽可能地简单。

10.9 知识集成与创新

在知识经济时代，创新尤其是自主创新是企业生存和发展的关键因素。自主创新包括原始创新、集成创新和消化吸收再创新等多种形式。无论采用哪种形式，创新过程都离不开对知识的运用。实际上，创新过程是一个包含了多种知识活动（如知识转化、生成、学习和重用等）的十分复杂的过程。由于知识创新涉及的因素较多，限于篇幅这里不做过多的论述。下面主要针对创新中的知识集成活动进行描述，其他活动已在本章前面介绍过了。

10.9.1 知识集成

知识的集成是对分散在企业中、存在于员工头脑中的隐性知识和分散在企业外部的知识进行集成。知识集成包括企业外部知识集成和企业内部知识集成两种。企业外部知识集成是企业吸收外部知识并与企业积累的知识结合而产生新知识的过程，而企业内部知识集成就是将存在于企业员工头脑中的个体的隐性知识集成为企业知识。

1. 知识集成的方式

知识集成主要有如下四种不同的方式。

(1) 隐性知识之间的集成，即知识在员工之间的转移、共享和集成。企业的创新在很大程度上依靠隐性知识之间的集成，关键是如何使员工积极主动地贡献自己的知识让企业集体共享，这里需要能促进知识集成的企业文化、组织管理机制和集成工具。基于互联网/内联网的多媒体信息技术将成为集成的基本技术。通过网络的交流，使私有的知识变成共享的知识，使支离破碎的知识变成完整、独创的知识，并通过知识的交流和集成产生出新的知识。

(2) 编码型知识之间的集成。随着互联网的普及，编码型知识呈爆炸态势。从浩淼

的知识海洋中快速找出企业所需的知识，需要有效的知识挖掘工具，当然更需要进行编码型知识的有序化。

(3) 从隐性知识到编码型知识的集成，即将人们头脑中的经验和诀窍总结出来，并进行集成。企业要取得成功，必须很好地开发、积蓄和利用自己的知识资源，为企业创造更多的价值。例如，销售人员在销售中的一些成功经验和重要发现应及时加以总结，并在企业内其他销售人员中推广。这需要有一种奖励机制和知识提取（发现）工具。

(4) 从编码型知识到隐性知识的集成，即通过培训和先进的知识传播工具将书本知识转化为人们头脑中的知识，因为创新最后要通过人进行。多媒体技术和虚拟现实技术可以提供这方面的支持。

2. 知识集成模型

知识集成模型的用途是描述知识与知识、知识与人、知识与过程的关系。知识集成模型处于发展之中，有以下值得研究的内容：

(1) 知识基元的分类系统。该分类系统对知识进行有序化，建立知识基元，并进行分类。

(2) 知识结构图。知识结构图定义相关的知识以及将这些知识和信息结构化，为知识的共享提供必要的条件。

(3) 知识图。知识图对组织中谁拥有何种知识进行描述，这对知识的共享和利用有很大好处。

(4) 知识需求和知识建立的过程模型。该过程模型定义某一特定的活动需要哪些知识，哪种知识是通过特定的文档化活动提供的，等等。

3. 知识集成工具

使用知识集成工具的目的是将知识与知识、知识与人、知识与过程集成起来，将知识与特定过程和未知情况进行动态匹配，以促进企业的创新。知识集成工具的以下内容值得研究：

(1) 知识库管理系统。该系统应能有效地对各种编码型知识进行管理。

(2) 知识提取（发现）工具。该工具应能通过群体协作、过滤和语义技术将隐性知识转化为编码型知识。

(3) 知识挖掘工具。该工具应能从数据库中挖掘出有用的知识。

(4) 知识集成平台。该平台应能利用多媒体技术，通过多种表达方式方便地进行知识交流。互联网的迅速发展为知识集成的实施提供了较理想的条件。

10.9.2 知识创新

美国麻省著名的恩图维国际咨询公司总裁、著名战略研究专家戴布拉·艾米顿(Debra

M·Amidon)最早提出了知识创新的概念,他将知识创新定义为“为了企业的成功,国家经济的活力和社会进步,创造、演化、交换和应用新思想,使其转变成市场化的商品和服务的活动”。

知识创新是一个系统的行为,在企业知识创新过程中,企业知识存量是其创新的前提条件,企业知识越丰富、越专业,内、外部环境越适宜,其知识创新活动的数量和质量就越高,知识创新的能力就越高;而企业只有在适宜的内、外部环境中将企业知识加以应用才能实施知识创新活动,知识创新的能力才能越高,其知识创新活动所带来的收益就越高;企业获得知识创新的收益必然一方面使得自身竞争优势加强,另一方面使得企业内外部环境进一步优化,这样加强后的企业竞争优势以及优化的内、外部环境又会通过提高企业知识创新能力使企业获得更大的收益,从而形成一个知识创新的循环过程。

知识创新的动力来源于企业外部环境的动力和源于企业自身的内部动力。企业外部环境动力主要包括科技发展水平、政府政策和法律制度、总体经济环境、市场需求、相关产业因素、社会文化和教育发展及人才培育这几个方面;企业自身内部动力则包括企业利益驱动、知识的愿景规划、组织管理、人才资源、财力资源以及企业家精神等。



思考与讨论题

- (1) 请分析知识系统的组成要素及主要功能。
- (2) 知识系统与知识管理系统是否相同?如不相同,它们的差别是什么?
- (3) 知识系统中的各种活动是相互独立的,还是有机地联系在一起?为什么?



参考文献

- [1] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [2] 杨小云, 陈雅. 知识需求与提供研究[J]. 情报杂志, 2004, 3: 89-91.
- [3] 陈菊红, 赵斌. 准时知识管理及知识供需的匹配分析[J]. 研究与发展管理, 2005, 17(5): 51-57.
- [4] 吴晓波, 彭新敏, 丁树全. 中国企业外部知识源搜索策略的影响因素[J]. 科学学研究, 2008, 26(2): 364-372.
- [5] Nonaka I., Takeuchi H. The Knowledge-creating Company[M]. Oxford England: Oxford University Press, 1995.
- [6] 程志. 知识系统的知识重用方法[J]. 计算机技术与发展, 2006, 16(7): 121-122.
- [7] 王玉梅, 张靖. 基于系统动力学的组织知识创新影响因素分析[J]. 青岛科技大学学报(社会科学版), 2009, 25(4): 58-63.

第 11 章

知识管理的典型应用

内容提要

知识经济是知识管理产生的时代背景，企业知识管理是知识经济时代企业的必然选择。知识管理除了在企业层面得到深入研究并进行成功应用外，还广泛应用于个人层面、项目和团队层面以及政府层面。本章首先介绍个人知识管理的含义、实施与技术支持工具，然后对项目管理中的知识管理进行讨论，其次介绍企业知识管理的含义、实施步骤以及典型的企业知识管理系统体系结构，最后介绍政府知识管理的含义、实施条件和政府知识管理体系。

本章重点

- 个人知识管理
- 项目知识管理
- 企业知识管理
- 政府知识管理



11.1 个人知识管理

不但企业需要通过知识管理来提高生产和管理水平,推动自主创新以产生新的价值,其实个人也需要通过知识管理来提升自己的工作和学习的能力,从而得到更全面的发展。

11.1.1 个人知识管理的含义

个人知识管理就是对个人的知识进行获取、组织、存储、利用和创新的管理,管理过程是在不断明确自己的知识需求的基础上,有效地识别和获取、整理和存储、集成和开发自己知识的过程^[1,2]。

有的学者认为,知识管理是从个人开始的。在这个基础上建立团队和组织的知识管理,然后再扩展为组织之间或更大范围的知识管理。事实确实如此,无论是个体知识工作者(如理论研究人员、律师和作家等),还是团队成员,都必须从明确自己的知识需求开始,然后收集、整理和运用这些知识,也可能在此基础上创造新的知识。可以说个人的知识管理乃是组织知识管理的基础^[3,4]。

个人知识管理的主要方式之一是学习。在科学技术快速发展的形势下,不学习就很难适应工作岗位的要求。个人的学习不仅限于在校学习或培训,在实践中通过干中学、在生活中耳濡目染不经意的学习也都是学习的方式。个人的学习主要是在自己的意愿支配下进行的,组织不一定会特意安排。组织对个人的工作要求已经蕴含着对知识、对学习的要求了。

个人学习不仅要掌握知识,而且要培养能力和创造力。掌握知识是人人都自觉认识到的,但是对于能力的培养却不那么容易自觉认识到。知识和能力既有联系又有差别。能力是在掌握和运用知识的过程中逐步形成和发展起来的,有一些能力就是学到并加以应用的隐性知识。但是能力不像知识那样一直在积累和丰富,而是在一生的某一阶段成长得比较快,而后就发展得缓慢起来甚至会减退。能力的发展要比知识的获得困难得多,现在有许多高分低能的青年,正说明了忽视能力培养的现象还大量存在,不能不引起注意。

个人学习可以增强自身的竞争力,而组织的竞争力也是在个人的优势基础上经过团队的合作逐步形成的。无论是个人学习还是组织学习,其成果就是知识的获得。获得的知识需要加以管理。实际上个人的知识管理在很久以前人们就已经不自觉地进行了。只是知识经济的发展使得人们对知识越发重视,自觉的个人知识管理也就提到日程上来了。

个人知识管理不仅是就业或创业的人才需要,在校的学生其实也需要个人知识管理,只是与知识工作者的目标和要求不同而已。

11.1.2 个人知识管理的实施

1. 个人知识管理的方法与步骤

要进行个人知识管理，首先要制定知识管理的战略，也就是明确自己需要和运用知识的最终目的是什么，为自己树立一个目标。根据这一目标，针对组织和环境对自己的要求，分析自己的优势和劣势，找出自己最需要建立的知识体系，以及获取的轻重缓急，对具体的知识管理定下方向。然后就可以按照一定的步骤进行管理的实施^[4]。

一般说来，可以按照下面的步骤进行：

(1) 对个人的知识需求进行分析。每一位知识工作者的知识需求一方面决定于他的岗位需求，另一方面决定于他的职业规范、兴趣爱好以及环境需要等。还要考虑到组织中的分工协作、人际关系等。除了考虑当前需要之外，还得考虑今后进一步发展的需要。这些需要最好能够罗列出来。

(2) 对所需知识进行收集。在明确了知识需求之后，需要从各方面收集知识。知识可以通过直接方式（如亲自调查、当面请教和进行实验等）和间接方式（听讲、阅读和查询等）获得。现在能够获得知识的渠道很多，比较重要的第一是通过学习，包括正规学习、继续学习和培训等。对学生来说，这是主要获取知识的渠道。对于在职的知识工作者来说，现在提倡终身学习，各种机会还是很多的。第二是从各种书籍、报刊、文档、新闻媒体中学习，这是过去的主要知识来源，现在仍旧是主要来源之一。第三是从互联网上获取知识。互联网是一个庞大的知识库，可以获得各种类型的知识，特别是最新的知识。第四是通过人际关系获得前面三个方面难以获得、甚至于是根本无法获得的知识，其中最主要的是工作经验、个人体悟的知识，因为这些隐性知识只有通过人与人的接触才可能外化。有心人有时也能从一些道听途说、街谈巷议中获得一些知识线索。还有就是从组织的知识库中直接获得知识，这是最便利的方法。

(3) 进行知识的组织与存储。人们不可能一次性地获得全部所需的知识，在不同时间内获得的零星片段的知识需要组织整理，使其归类或者系统化。其实还有一个过滤的问题，去掉重复的、无用的知识。散在各处的、在不同时期得到的知识经过组织整理，还要采取有效的方式和工具加以存储备用。这时需要选择适当的个人知识管理支持工具，本书将在 11.1.3 节中详细探讨这一问题。

(4) 进行知识集成和新知识的获得。零星片段的知识经过整理后已经可以方便地取用，但是针对使用者特定的需要，还必须挑选出所需的知识加以集成，形成整体。另外在解决问题的过程中，当现有的知识不足时，还得通过知识的发现和创新，获得新知识。尽管知识发现和新知识的生成离不开群体，但是真正实现还是在于知识工作者个人。

(5) 进行知识交流与共享。在个人知识管理的基础上可以进行知识的交流，做到知

识共享。由于个人知识已经得到整理，所以传播交流都很方便。

(6) 对知识管理不断进行评估。需要不断结合自己的工作和学习，进行个人知识管理的评估，发现不足之处及时纠正和改进，过时的和无用的知识要及时清除。

2. 个人知识管理需要的技巧

文献[5]指出，进行个人知识管理需要下列七种技巧：

(1) 检索信息的技巧。在个人知识管理中，检索信息的技巧既包括技术要求很低的向他人提问题、听回答的技巧，也包括充分利用互联网的搜索引擎、电子图书馆的数据库和其他相关数据库查找信息的技巧。为充分掌握检索信息的技巧，个人有必要对搜索的概念、搜索的算法、搜索的技能等有一定的了解。

(2) 评估信息的技巧。这种技巧不是仅指可以判断信息的质量，而且还指必须能判断这种信息与自己遇到的问题的相关程度。在评估信息和知识的时候，个人没有必要去了解计算机评估信息的机理，而是将评估的准则放在可信度、准确度及相关支持等方面。可信度一般根据对作者的置信度、质量保证依据、元信息等来判定；准确度可从时间界限、综合性、信息面向的对象及其使用的目的性、合理性等方面来确定；相关支持则是指信息文本的索引目录、参考文献等。

(3) 组织信息的技巧。组织信息，需要过滤无用和相关度不大的信息资源，有效地存储信息，建立信息之间的联系，以便于以后的查找和使用。有效组织信息的原则是：无论环境怎样，组织起来的信息应该便于有效的利用。这种技巧会牵涉到用不同的工具把各种信息组织起来。在手工操作的环境中，我们会用文件夹、抽屉和其他比较原始的方法来组织信息。在现代高科技环境中，可以用电子文档、数据库和网页，或者用专门的知识管理软件来组织信息。

(4) 分析信息的技巧。分析信息就必须牵涉到如何对数据或文本进行分析并从中得出有用的结论。常用的分析信息的方法是建立和应用模型，通过大量的数据分析从而得出信息间的关系。电子表格、统计软件、数据挖掘与文本挖掘软件等提供了分析信息的方法，但在建立各种使用分析软件模型的工作中，人的作用还是最重要的。

(5) 表达信息的技巧。通过表达信息，可以实现隐性知识向显性知识的转化，使个人知识在交流、共享中得到升华。信息的表达，无论是通过屏幕图形演示、网站，还是通过文本，大部分的工作应该围绕如何让其他人理解、记忆，并能与自己进行互动来进行。可视化技术的应用是很重要的。

(6) 保证信息安全的技巧。虽然保证信息安全的技巧与个人知识管理中其他的六种技巧有所不同，但这并不表明保证信息的安全就不重要。保证信息的安全涉及到开发与应用各种保证信息的秘密、质量和安全存储的方法和技巧。常用到的密码管理、备份、

档案管理都是保证信息安全常用的方法。

(7) 信息协同的技巧。信息技术的发展为组织和部门的协同工作提供了强有力的支持。如通过小组或团队的形式组织交流和学习,领导、专家和一般成员在讨论与交流的基础上可以对一些要解决的问题进行协同工作,交流和共享彼此的观点和知识。有效的利用这种技术不仅要求会使用有关的方法,而且要求充分理解协同工作的各种原则及其内容。

11.1.3 个人知识管理的支持工具

对于知识工作者个人来说,有一些可以帮助提高工作效率和效能的工具(计算机软件),下面依次介绍个人生产率工具、信息网络工具和电子绩效支持系统这几类有关的信息工具^[4]。

1. 个人生产率工具

个人生产率工具主要用于文牍工作。例如,中国的金山 WPS、美国微软公司的 Office 等,就是这类成套的软件。其中主要的部件包括:

文字处理工具。它是用来以电子文档形式对文件进行编写、修改、复制、传递和保存的。这种电子文档便于编写、维护、传送和使用,如微软的 Word 就是这类工具。它的处理包括:文档录入、文档加工和文档输出。在中国,有一个特殊的问题,就是汉字的输入/输出问题,目前这一问题已经解决得很好。

电子表格。它是以表格形式记录、汇总和处理大量数据,代替手工绘制、填写表格和进行统计、计算、制图等工作的工具。Office 中的 Excel 就是这种工具,金山的 WPS 中也有这一功能。它的处理包括:表格的建立和录入、表格的处理、表格的输出。电子表格是一种非常有效的数据记录和处理工具,其在国外财务、工程、物流等方面使用很广。过去在中国的使用还不普及,因而人们对它的潜力认识不足,随着数据处理工作量的日益增加,它的应用正逐步扩大。

报告演示。在进行各种技术或公务报告时,常常需要一些图表、文字资料为与会者进行演示,如果使用纸面图表或者幻灯片,制作是很费事的。现在有了报告演示软件,如 Office 中的 Power Point,就可以事先在计算机中制备好,在演示时通过投影机放映。由于软件提供了非常方便的文字、图形和图表的输入方式,特别是可以使用现成的图形、图像和视频资料,演示鲜明生动,因而使编写演示文稿变得轻而易举,效果是非常好。

数据管理。对于知识工作者个人来说,经常有一些数据需要管理,尽管在一般的文字文档中可以放置数据,但不利于检索处理。使用电子表格也可以保存和处理数据,但当数据量大时也显得不方便。所以在像 Office 这样的成套软件中,也有像 Access 这样的

小型数据库及其管理的功能。用户可以在个人范围内很方便地使用它，并可通过像 ODBC 这样的功能与大型数据库进行相互转换。

工作管理。这是为知识工作者个人提供的计划、日程安排、工作日志和备忘录等，以及用于与其他人员联系的工具。例如，Office 中的 Outlook 软件以及一些专门的软件。这部分工具的进一步发展就是计算机支持协同工作（CSCW）系统了，这在第 9 章曾经讲到过。

在使用上述工具时，为了拟定文档格式或者编制表格需要花一些工夫，等到下一次做同样工作时，又得另起炉灶，显得十分烦琐。所以在文字处理、电子表格、报告演示的制作过程中，可以应用模板。所谓模板就是把经常使用的文件、表格、演示图形等的共同部分，预先做成标准格式。人们在建立文件、表格和演示文件时，选择相应的模板，在模板的基础上继续建立内容与格式。这样就可以省掉许多重复性的劳动。在一些常用的工具软件中，如 WPS、Word 和 Excel 之中就有通用的模板可以选用，略微专门一点的，用户可以自己先编制的一些备用。

2. 信息网络工具

随着互联网和内联网的普及应用，知识工作者经常要上网，对于信息网络工具也要有所掌握，这包括：电子邮件、万维网浏览与文件下载、内联网浏览与文件交换、博客、维基等工具。

由于网页作为表达、保存、传递信息与知识的最为方便的单元，通过超级链接又很容易建立信息与知识的关联，所以许多知识工作者就用网页形式来保存知识。特别是一般工作人员认为使用数据库的结构化要求太高，不如网页方便，因而不拘形式地把自己的知识库以一堆网页的集合方式逐步堆积起来。

电子邮件是当前通过互联网进行通信的最为普及也是最为经济的方式。它可以传递文字、图形、图像和语音等信息，使用方便，投递迅速，全球畅通，而费用几乎低到可以忽略不计。只要用户有一个电子邮件信箱，向相应的服务网站登记注册，连通后即可收/发信件。

当前许多企业在互联网上进行信息活动，主要的形式之一就是电子邮件。收/发电子邮件几乎成了每日必不可少的活动。从知识管理的角度来看，电子邮件是进行知识交流的最常用的工具。在电子邮件的数量多起来以后，邮件的管理问题也开始提上日程。现在的电子邮件系统，如微软的 Outlook Express，中国的 Foxmail，都可以通过建立文件夹进行分类存储，并有一定的按发信人、按日期等进行检索的功能。

正因为电子邮件的内容有不少是与知识的传递有关，所以人们在想：怎样把个人的知识库或者组织记忆与电子邮件系统联系起来，不论是对接收邮件进行归档，还是选择

有关的知识发送出去,都能使邮件与知识存储挂钩。这里存在一个怎样对组织记忆或知识库进行与邮件相对应的检索问题。现在有一种知识强化的电子邮件系统,英文简称 K-mail。它的基本思想是在电子邮件系统和组织记忆或知识库之间,加上一个中间索引环节。由于电子邮件像其他知识文档一样,除了内容之外,还有相应的语境或者说背景问题。电子邮件的语境包括形式化的发信日期、发信人姓名、单位、职务、发信的邮箱地址等,但还有与信的内容有关的语境。怎样从信件内容或者组织记忆中提取这些语境信息便是有待解决的问题。

3. 电子绩效支持系统

对知识工作者进行支持的工具,还有一种叫做电子绩效支持系统(EPSS, Electronic Performance Support System),它是近年来出现的一种新的信息系统。

美国的Gery是最初提出这一概念的^[6]。按照他的定义,电子绩效支持系统是一种可以按照需要,从整体上提供信息、建议、学习经验和工具,使人能在最少需要他人支持的条件下得到高水平的绩效。后来有人把这一定义加以扩展,认为电子绩效支持系统是一种电子基础设施,用来在一个组织的范围内获取、存储和发布个人与组织的知识资产,使个人能以最短的时间以及得到最少的他人支持力度达到所需的绩效水平。

还有人认为,电子绩效支持系统是一种集成型的绩效支持系统,它不仅是一种技术,而且是为了满足组织的当前与未来绩效需求的战略。也有人认为它应该是动态的绩效支持系统,它的能力会随着经验的改变,会由使用者加以更新和调整。

电子绩效支持系统的目标是:无论在什么情况下,一旦有需要就能产生绩效与适宜学习,因此这类系统被认为是“即时训练”(JIT Training)工具。对新手来说第一天工作使用它时就能产生绩效。不但支持今天的工作,而且可以帮助建立未来工作的知识基础设施。

电子绩效支持系统中的知识表达方式是多种多样的,可以是文本、数据、图形、图像和视频等形式。电子绩效支持系统的主要组成部分有:人-机接口、工具、信息库、建议系统和学习经验系统等。

电子绩效支持系统的开发可以采取渐进的方式,先按照某一系统的前端开发,或者开发成为现有系统的一个补充部分,再开发针对某一特定任务的独立系统。最后再发展成集成型的系统。

由于电子绩效支持不仅是一种技术,而且可以看作是一种战略,后来有人把它与组织学习结合起来,便形成一种新的工作模式,即电子绩效支持下的组织学习与系统开发。

本书第9章曾经介绍过的博客和维基也都是容易使用的个人知识管理系统。利用博客和维基进行知识管理,作用在于知识收集、知识提取、知识共享和知识应用等过程。

还可以用来与他人进行知识共享和交流，在知识管理领域有着广阔的应用前景。

11.2 项目管理中的知识管理

本节主要探讨项目管理中如何组织和管理知识，使其在项目进程中发挥应有的作用。

11.2.1 项目管理的特点

当前，项目管理作为一种科学的管理方式，日益受到人们的重视。这是因为大量的工程建设、研究开发工作和社会事业的兴办都是按照项目来加以组织和管理。即使是企业内部的日常管理，由于当前环境的急剧变化，按照常规办事常常满足不了要求，也得把一些任务按照项目来筹划、组织和运作。因此项目管理开始受到前所未有的重视。

项目的含义很广，从古代的长城、都江堰建设到现代的人类基因组计划、登月计划，都可以看作是项目。一般说来，所谓项目，就是一项有特定指标的专门任务，需要在一定的资源条件下，由一定的组织在计划的时间内一次性完成。从下面列举的项目特性，可以更具体地了解什么是项目。

(1) 项目的一次性。这是项目和经常性任务（如工业企业中的正常生产）的最大区别。项目有明确的开始时间和完成时间，一次完成，不会有完全的重复。

(2) 项目的独特性。每个项目都有与其他项目不同的独特之处，有些项目可能沿用了过去类似的设计或技术，但一定有它明显不同之处，更不用说一些创新项目的开创性特点了。

(3) 项目的多目标性。项目兴办的目标常常不是单一的，它的具体目标属性，如功能、影响、成本和时间等也是多方面的。这些目标有时是协调的，有时候是相互矛盾而需要协调的。

(4) 项目组织的临时性。一般都是为项目专门组织队伍，完成后解散，下一个项目再重新组织。

(5) 项目寿命期的分阶段性。从项目的选定到完成、评估是分成若干阶段的。

任何项目都会涉及多方面的因素，需要很好地加以组织管理。所谓项目管理，是指在项目的活动中，运用行之有效的理论、方法、技术和工具来计划、组织、指挥、控制和协调项目生命期中各阶段的工作，以达到预定要求的过程和活动。项目管理所涉及的要素主要有：

(1) 资源。这是项目实施的基本保证。它包括自然资源与人工资源、内部资源与外部资源、有形资源与无形资源。由于项目的一次性特点，对于资源的获取、使用和管理与一般生产有很多不同的地方。这里特别要提到的是知识作为一种无形资源，在项目管

理中有着特别重要的意义。

(2) 需求和目标。由于项目的利益相关者（包括客户、供应商、项目投资方、执行者和其他利益相关者）有着各不相同的要求，有的甚至相互矛盾冲突，需要加以协调，确定适当的目标。

(3) 项目的组织。项目的组织包括机构、领导、章程、人员配备和运行机制等。由于项目的一次性与阶段性特点，组织总是在不断变化的。

(4) 项目环境。项目环境包括政治和经济环境，文化和思想意识，以及标准和规章制度等。

项目管理的基本职能有：① 项目的计划：根据项目的目标要求，对各项活动做出合理的安排；② 项目的组织：这包括两个方面，一是组织机构，二是组织行为；③ 项目的控制：对项目按照计划进行并随时加以监督和必要的调整；④ 项目的评价：在项目进行过程中及时发现偏差并加以调整，保证计划的实现；项目完成后对项目的成败得失做出恰当的总结和评价。

随着中国社会经济的高速发展，各种建设事业和研究开发工作在不同行业、不同地区蓬勃开展，项目的数量和规模日益增大。但是相当一部分项目的组织和管理存在着混乱现象，使得资金、人员、质量和进度等方面发生严重失控现象，各种拖期、超支、浪费和侵吞国家财产情况屡有发生，造成了大量的损失。因此运用现代化的项目管理方式来提高项目的组织管理水平、保证建设事业的顺利进行，是当前极为重要的任务。

11.2.2 项目管理中的知识

项目管理作为一种科学管理方式，需要建立它的学科体系。近年来世界各国的项目管理专业组织纷纷建立各自的项目管理知识体系（PMBOK, Project Management Body Of Knowledge），作为项目管理专业的知识基础。中国也进行了相应的工作^[7]。

全面理解和建立科学的知识体系，将能对知识在项目管理各个方面、各个阶段的作用有更加全面和深入的认识，对在项目管理的进程中如何有效地实施知识管理有一个比较切合实际的构想。项目管理知识体系中包含的主要是显性知识。这些知识体系提供了交流中的共同语言和进一步发展的知识基础。

美国管理项目协会（PMI）1996 年颁布的项目管理知识体系大纲划分了 9 个知识领域，包括项目整体管理、项目范围管理、项目时间管理、项目成本管理、项目质量管理、项目人力资源管理、项目沟通管理、项目风险管理和项目采购管理。各个领域均有相应的知识内容。

国际项目管理学会（IPMA）也编制了自己的项目管理知识体系标准，称为《国际项目管理专业资质标准》（IPMA Competence Baseline），其特点为：对项目管理人员提出

40 个方面的素质要求（核心要求 28 个），强调项目管理人员的实践背景，注意与专业技术知识的结合，允许成员组织按照民族、文化以及职业发展要求，变更非核心要求。

在中国，中国优选法统筹法与经济数学研究会项目管理研究委员会发起组织了中国项目管理知识体系的研究制定工作，推出了《中国项目管理知识体系》（C-PMBOK）^[8]。它包含项目与项目管理、公用知识、方法与工具三方面内容，各个方面包含了相应的知识模块，共 88 个模块。

实际上，项目管理知识体系的知识范畴应该包括三个部分：项目管理特有的知识，一般管理的知识，与项目有关的应用领域的知识。上面列举的知识体系实际上只包括第一部分。

这些知识体系的制定与发布，对于促进项目管理专业化发展和提高项目管理人员的水平，都将起到积极作用。对于项目管理人员的培训教育，对于项目管理人员的资质认证，也将提供参考。

但是在项目管理的实践中，仅仅具有显性知识是不够的。因为这些知识提供的是一般的原理，在应用于具体实践中的时候，还要根据具体的主、客观条件加以选择、综合和灵活处置。特别是项目管理的一次性特点和独特性特点，使得它与一般的运作管理相比较，更不能一切都按照成规办事而需要有独立的抉择和处置方式。这时候，包含经验、直觉、洞察力在内的隐性知识，将发挥重要的作用。由于它包含着丰富的发现问题与研究问题的直觉和洞察力，因此如果能使它充分发挥作用，将能使问题的解决、新思路的形成更加迅速有效。

这里说的隐性知识，不仅包括个人的隐性知识，而且包括组织（班组、团队和企业）的隐性知识。这些隐性知识包含在组织的共同信念和目标、工作反思和工作习惯之中，是组织的核心竞争力。

正因为如此，我们在谈到如何组织管理知识以推动项目管理工作时，必须同时考虑各类知识。只注意显性知识而忽视隐性知识是难以推动工作的，但只注意隐性知识而忽视显性知识（这种情况发生在一些全凭经验工作的中小型工程项目管理中）就会使项目管理无法建立在科学的基础之上，即使一时取得成功，也难以保持长久。

在项目管理过程中应用和管理知识，与在其他领域中一样，首先要明确的问题包括：究竟需要什么知识？这些知识在哪里？显性知识存放在哪里而隐性知识又是由什么人所具备？为使各类知识都能有效地发挥作用，应该创建什么基础设施（如信息技术网络）和营造什么文化氛围？

11.2.3 项目各阶段的知识管理

在项目生命期的各段时间里，需要和使用的知识也是各具特点的。由于不同性质的

项目其阶段的划分也不各相同（如建筑工程项目与科研项目的阶段划分就有所不同），所以这里只按项目生命期中的各项任务来分别讨论。

文献[9]提出可按下面六个阶段来研究知识的运用：① 项目概念的形成；② 项目环境的分析；③ 目标和范围的确定；④ 项目的详细规划，资源配置，项目的批准；⑤ 项目的实施、控制；⑥ 项目的评价和推广。

实际上这六个阶段有可能重叠、交叉，甚至出现反复，不同阶段知识的运用也可能有所跨越，但都要进行下列知识活动：① 寻找、发现知识源；② 获取显性知识，如文件和软件等；③ 组织沟通以共享隐性知识。

在形成概念的阶段，显性知识还不能很好地组织起来，人们运用的还是隐性知识与人际关系，但这些却形成了项目的一般概念，构成了项目有关各方共享的知识基础和日后知识工作的起点。这个阶段可能产生的知识，是有关项目最初的但是极有价值的知识。一般来说，项目概念形成阶段的知识管理工作还得不到正式的人力和预算的安排，可能只是少数人在进行，但还得进行下面的工作：① 从现场收集与项目有关的文件与信息，最好建库保存；② 将所有与项目有关的讨论与沟通信息记录下来，从而获取隐性知识；③ 建立初始的数据库；④ 建立有关的名词术语目录；⑤ 向项目相关者解释知识管理与建立共享知识的文化氛围的重要性。

在对项目环境进行分析的任务中，需要辨认对项目有影响的所有因素，因而需要下列诸方面的知识：① 与项目有关的经济、政治、法律、社会和地理知识；② 专业知识（如水资源、卫生等）；③ 战略信息（当地权威机构的支持、合作伙伴情况等）；④ 组织管理方面的知识（领导隶属关系、财政金融等）。

这些知识可能分散在各处，可以通过建立知识路线图的办法来进行知识的组织。

在目标和范围的确定、可行性分析与风险分析工作中，专家的隐性知识将起到主导作用。这是建立共识和决策的过程，一方面需要大量显性知识，另一方面迫切需要直觉和现场经验。这时需要建立（电子或纸面的）文档库，以及有关专家的信息（隐性知识的源泉）。

现在需要研究的问题是：项目相关者的期望是什么？对项目范围的约束是什么？有什么可以预见的风险？时间与成本如何平衡？

在这一阶段，除了电子的或纸面的文档库外，应该特别注意建立隐性知识的交流和储备手段，甚至建立网站以利于人际交往。这时正是建立知识管理系统的最好时机。

在项目的详细规划，资源配置，项目的批准过程中，由于要进行工作分解、进度安排与资源配置，使用的多半是显性知识，需要使用相应的软件，如知识管理系统（显性知识的外化），但仍旧需要保持隐性知识来源（如专家联系表）。

项目的实施、控制过程是考验前几步的关键阶段，这时知识管理的任务一方面要及

时提供所需要的知识，并检验其正确、恰当与否；另一方面要收集、记录项目进行过程中产生的知识，并时时改进知识管理系统。这时的知识管理系统，联系了方方面面，特别需要把显性知识与隐性知识进行有机的集成。任何有关的会议、文件、广播电视报道常常能提供有用的知识。这时会有许多新知识的生成，应该不失时机地把它收集起来，否则时过境迁，这些宝贵的知识财富就会流失掉。

从知识管理的角度来看，项目的控制是一个学习的过程。不但要从报表、文件学习，还需要从有关人员处学习。新的知识需要及时输入知识管理系统。有些信息与知识要能使项目相关者很容易就能获取到。一些决策支持工具（如统计、评价和方案选择等）在这里是有用的。

在项目评价与推广阶段，应该特别强调学习的功能，因为这时已经可以总结项目的经验和教训，并把结果（包括项目本身的成果）推广出去。这时除了已见诸文字的显性知识外，还应该利用各种方式收集隐性知识，使其转化为可以传播的知识。一个优秀的项目管理团队，应该是一个学习型组织。

现在，由于微型计算机及其网络的推广应用，在项目管理中使用了不少的计算机工具，特别是财务与工程进度管理（如网络计划）系统以及计算机辅助设计系统等，大大提高了管理的知识含量。而又由于多媒体技术的广泛应用，一些直观的信息手段使得人们能远距离了解工程的形象进度，为现场管理和隐性知识的交流提供了有力的手段。

在项目管理过程中，不但要使用现有的知识，而且在解决新的问题时还会创造新的知识。

特别是由于项目的一次性和独特性特点，每个项目的进行过程中，总会有一些新知识产生；而由于一次性的特点，项目结束后组织解散和人员变动会使这些知识流失。所以特别要注意收集。

在项目管理过程中，需要进行一系列的决策，这些决策涉及的范围可能不同，但只要不是按一些例行的成规进行抉择，每次都会面临新形势，都可以认为带有不同程度的创新，这就更需要把显性知识与隐性知识融合起来应用，并在决策过程中获得新的显性和隐性知识。

不论是组织还是个人，不要因为项目的一次性特点而忽视了知识的累积。这种累积一方面表现为文档资料的收集整理，另一方面表现为人员或组织的隐性知识（经验、直觉判断力等）的增长。这种增长常常是不自觉的，但如果人们自觉地去总结和反思，会有更大的收获。

11.2.4 专家知识的获取

下面再讨论一下怎样从具备隐性知识的人员那里获取知识的问题，特别是当他们分

布在不同地点的时候^[3,4]。

由于项目的一次性与独特性的特点，常常需要一些专门领域中的经验知识，这些知识保存在一些专家的头脑里面，而没有形成文字材料（因为具备这些知识的人在完成一个项目之后接着参加下一个项目，来不及总结，再加上有些经验知识本来就是无法言传的）。这些专家由于分处各地，可能属于不同的部门和专业，共同语言、甚至文化背景都不一样，因此不具备面对面交流讨论的条件。同时，随着经济全球化的快速发展，跨国项目日益增多，上述问题就显得更加突出。

具备隐性知识的专家，并不一定自觉自愿地将他的知识贡献出来，所以获取这种知识要比获取显性知识困难得多。但是在许多项目的进程中，这些知识又是很重要甚至是很关键的，所以应该设法将这种获取很好地组织起来。

根据国外的经验，应该做好下面几件事。

首先是要把这件事列入项目的计划。要回答下面几个问题：① 需要什么专业知识？② 专业领域的特点是什么？③ 在每个领域中谁是专家？④ 专家在什么地方？⑤ 每位专家提供帮助的可能性有多大？

上述工作以及今后的联系都是由组织的知识中心来负责的。知识中心要先与有关专家建立关系，要以非正式的方式使专家了解项目的目标，他能够起到的作用以及他可能的贡献。一般来说，专家的工作仅限在一定范围之内，因此要给专家以一定的安排自由度。通常需要的仅是专家的一技之长，所以不能求全责备。在选定专家之前，要明确：① 专家的作用是什么？② 专家提供的是什么？③ 专家需要花多少时间和精力？④ 谁和专家一道工作？⑤ 工作计划日程是怎样的？⑥ 结果怎样使用？

在需求方面可能需要讨论几轮，然后达成共识。由于与有些专家过去并没有深入的交往，所以取得专家的信任是很必要的，这必须做到：① 对专家作用的尊重；② 为专家提供必要的工作条件和环境；③ 以邀请方的影响推动专家去完成工作；④ 通过有效的手段传播知识；⑤ 虽然可以使用远程会议工具，但必要时还得登门求教，面对面沟通。

此外，为了节省专家的时间和精力，有一些事务性、技术性的工作如文件编辑、打印和制图等，应由邀请方派人协助进行。

对于专家的工作，除了应该支付一定的物质报酬外，还应该注意对它的成果予以认定和宣传，使他有成就感。专家参加一个项目的工作并不仅仅是获得物质与精神报酬，有时候他也想通过这一工作扩大眼界，增长知识，通过成果得到认可，提高知名度。所以在工作中也要在不影响工作和泄露诀窍的条件下予以满足。

对于一个组织的知识中心来说，组织专家网络的能力也是一种竞争优势。这种组织能力表现为经常能提供专家名单和邀请到专家。这可以利用经常举办各种会议的方式来广为联系各个领域的专门人才。一些跨组织、跨行业的兴趣小组常常是专家的来源。

11.2.5 项目管理中的知识集成

项目管理过程是一个隐性知识与显性知识、个人知识与组织知识之间不断转化和不断创新的过程。一个项目的完成,除了该项目本身要实现的预定目标外,还包括项目信息档案等显性知识以及该项目运作过程中获得的经验教训等隐性知识的产生。集成这些知识,有助于企业今后相关项目的开发,有助于企业的长远发展。

项目管理中的知识集成的主要特点是:① 从项目组外集成项目管理所需的知识和信息,帮助项目组进行有效的管理;② 在项目组内集成项目管理所需的和新产生的知识和信息;③ 充分利用信息技术和知识集成技术,建立以知识和信息为基础的知识型组织和知识集成平台,促进知识和信息交流共享,培养项目组成员间的知识共享能力,创造知识和信息共享环境,提高项目组成员知识创新的能力;④ 将项目中积累的知识资源进行整理和规范化,以用于以后的类似项目,使项目管理知识得以继承和重用。

项目管理中知识集成的目标是:在项目管理中最大程度地获取、积累、传递、共享和利用知识,使每个员工在最大限度上贡献知识的同时,也能享用他人的知识,高效、优质地完成项目的管理任务。

在构建项目管理知识集成系统时,需要从项目管理运作各个阶段获取知识,然后通过系统分析和评价等知识处理过程,将知识重新运用到当前的项目管理运作中去,并不断充实企业的知识库,以利于今后项目的实施。

项目管理中的知识集成,不仅要在知识集成的运作层上,利用知识集成系统进行知识获取、知识传递、知识利用、知识交流创新以及知识评价,实现知识的社会化、外在化、综合化以及内在化的螺旋式处理;而且因项目管理自身的特点,项目组织的临时性、项目的一次性以及人员的流动性,项目管理中的知识集成更加困难,还必须结合项目组织文化,顺应企业的战略目标进行发展。

下面简要介绍一下文献[10]构建的项目管理知识集成系统功能模型。

图 11.1 描述了项目管理中的知识集成系统的主要功能,说明如何通过软件来支持项目管理中知识的获取、传递、利用、交流创新以及评价,较好地辅助项目管理中知识集成系统的实施。

知识集成系统中的知识获取模块主要从以下两个方面获取知识:一方面是对现有的一些显性知识的采集,如通过收集企业现有研究资料、客户资料、竞争对手资料,并创建电子文档输入到知识仓库;另一方面是对来自对隐性知识的挖掘和捕获,如项目开发部门或雇佣专家在内部创造和开发过程中获得的经验,用户反馈信息的分类、分析和提炼,然后形成对组织有价值的知识,并录入到知识仓库。

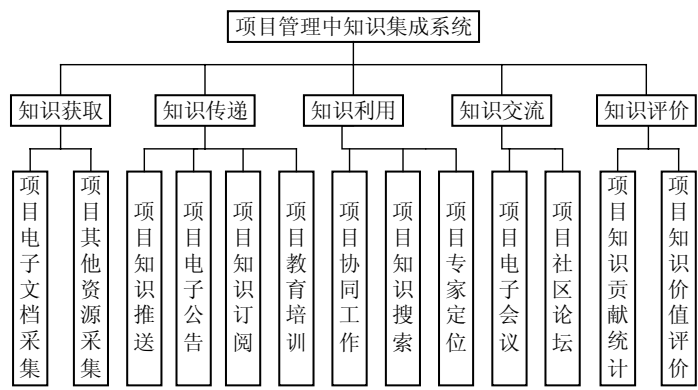


图 11.1 项目管理中的知识集成系统的主要功能

在知识应用前，知识应分配给需要知识的人。项目管理中的知识集成系统中的知识传递模块通过项目知识推送、项目电子公告、项目知识订阅、项目教育培训等主要功能子模块，利用页面浏览、电子邮件、文件传输等方式来实现知识的传递和分配。项目知识推送是指根据项目用户所从事的工作岗位、项目用户的兴趣以及项目的最新进展等情况，自动向项目用户推送相关的项目基础知识、技能和数据。项目电子公告是一种广播式的项目知识推送方式，具有及时更新、便于维护等特点。它不同于一般知识推送的地方在于其面向的范围更广，目的性不如一般知识推送明确。项目知识订阅是指项目用户可以按照自身兴趣、任务和工作重点等多种方式主动订阅知识，定期按时得到相关的最新知识，并且根据需要及时地更换订阅内容。项目教育培训是指项目员工可以通过选择所需培训的内容，选择网上教室，进行网上培训；开设课程的教师可根据员工网上作业的成绩和考试成绩等对参与该课程的学员进行评估。

知识利用是知识集成的目的。项目管理中知识集成系统中的知识利用模块提供了项目协同工作、项目知识搜索、项目专家定位等功能子模块，主要用于支持知识利用功能。项目协同工作功能主要是为处于异地的同一项目组成员对特定问题的研究和讨论开发的，如协同进行计划、设计和软件开发等。项目知识搜索是指搜索者根据许多特征，如排名、语言的选择、概要、关键词和日期等来定位搜索所需的知识，如各种技术文件、图纸、合同、规章制度、来往信函和传真、会议纪要、日常文件、工程进度、工程质量、工程成本以及项目运作中积累的各种经验和技巧等。项目专家定位是指项目运作过程中，在可能遇到各种各样的问题一时难以解决时，对拥有特定领域知识的专家进行准确、全面的查找和定位，可以方便用户与这一领域的专家直接进行交流，实现经验、技巧各种知识的共享，加快项目的进程。

项目管理中知识集成系统中的知识交流模块提供了项目电子会议、项目社区论坛等知识交流创新的良好场所，它可以帮助用户实现知识共享和知识创新。其中项目电子会议是指在存在时差和空间距离的情况下，通过召开电子会议来实现异地协作，最终使用

户可以很容易地获取所需的知识,方便与专家交流,及时有效地解决许多问题。通过积极有效的交流,参与者可以不断产生许多新想法和解决方案,实现知识的创新。在项目社区论坛中,用户可就某个共同感兴趣的主题进行交流和共享,可查询、阅读和以交互式方式访问所有成员开发的各类主题,并在多个社区之间自由切换,从而给论坛的各个发表者直接反馈。在论坛中,当用户产生新的想法时,可以开发一个新的讨论主题,这不仅便于其他用户的访问和讨论,也利于将这些论坛的内容为他人共享,更好地进行知识创新。

为更好地更新和发展企业的知识仓库,同时也为鼓励员工将自己的知识在项目内供他人共享,需要对知识进行正确有效的评价。知识评价模块为分析知识的利用率和创新评估提供了重要依据,同时也为激励员工贡献知识制定相应措施提供了量化指标。项目知识贡献统计功能包括系统自动评估发表文章的数量、质量等,以此来评估员工对知识库建设所做的贡献。项目知识价值评价功能包括在知识库的发展过程中对某一知识内部共享率的评价、在增进整个知识库的积累方面的评价以及对知识创新度的评价等。而对在讨论区发表的文章,可由社区成员进行评估,对特别优秀的文档,可交由专家进行评估。

11.3 企业知识管理

企业知识管理是近年来引起人们广泛关注的新兴学科领域,人们对此开展了广泛、深入的研究。

11.3.1 企业知识管理的含义

在知识成为主要竞争资本的今天,企业要维持竞争优势就需要依靠比竞争者更快的创造、传播和使用新知识的能力。随着现代信息技术和先进制造技术的不断发展、人才的频繁流动、竞争的日益加剧,实施以知识资源为管理对象的知识管理对于形成企业的核心竞争力至关重要。

企业知识理论认为决定企业能力和竞争优势的是企业拥有的知识和与知识密切相关的认知学习能力。知识管理有助于企业从现有数据中挖掘有用的知识,增强企业的商务智能,从而提高企业的认知学习能力和创新能力。

文献[11]提出了一个企业知识管理的综合概念模型,如图 11.2 所示。综合概念模型中,企业知识管理包含企业、团队和个人三个层面以及知识编码、员工发展和人际网络三个维度。

从个人层面来看,企业知识管理是从个人开始的。在知识编码维度,需要建立个人知识存储空间,为高效工作和及时存取知识提供便捷的手段;在员工发展维度,员工为了满足工作中的知识需求,应该不断充实自己的知识,构建并适时调整自己的知识结构

体系；在人际网络维度，员工应多与领域专家进行沟通、交流，通过人际关系获得专家的工作经验、个人体悟等方面的隐性知识。

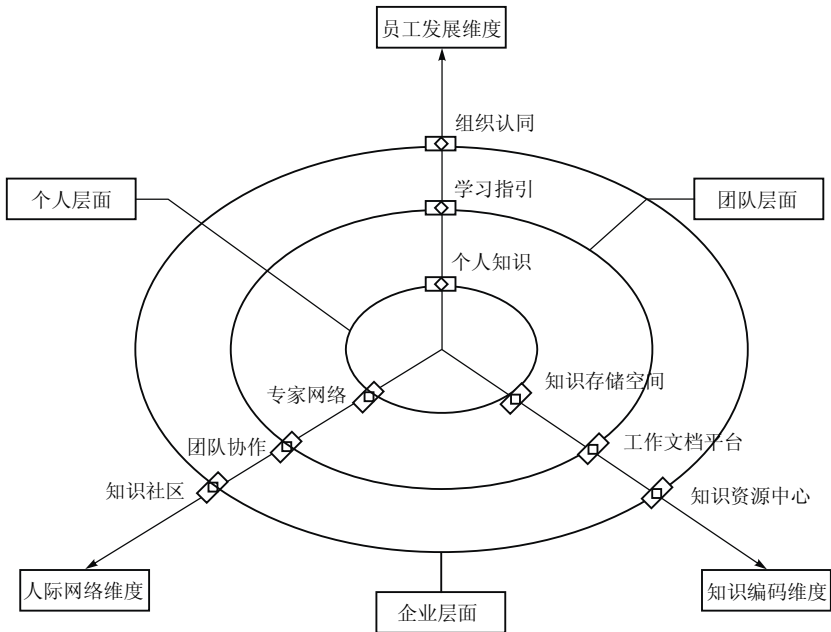


图 11.2 企业知识管理的综合概念模型

从团队层面来看，企业内部的日常管理和项目管理通常是基于团队来组织和运作的。在知识编码维度，需要在日常工作基础上积累和形成团队内部的工作文档库、技术资料库以及客户资料库，建立工作文档平台，供团队成员随时获取所需知识；在员工发展维度，应建立员工学习指引平台，帮助员工循序渐进地掌握领域内知识；在人际网络维度，鼓励团队协作，建立团队内部交流制度与和谐的人际关系，从而实现个性化策略的知识转移。

从企业层面来看，为实现企业的总体战略，应制定知识管理的愿景、使命、目标和战略措施。在知识编码维度，需要建立企业知识资源中心，建立、改造和运用与知识管理配套的科技成果管理流程、培训管理流程以及知识审核、发布和获取流程；在员工发展维度，应建立和运行一套具有基于知识贡献、共享、学习和创新水平的员工绩效考核体系和激励制度；在人际网络维度，建立知识社区和实践社群，实现员工随时网络化交流，倡导组织学习和共享价值观，创建学习型组织。

11.3.2 企业知识管理的实施

随着信息基础设施建设的深入，信息技术不仅成为企业运营的基础，而且在企业资源计划（ERP）、客户关系管理（CRM）、办公自动化（OA）等信息系统内沉淀了大量的知识，成为企业创新的知识源泉，于是知识管理逐渐提上了大中企业信息化建设的议事

日程。企业知识管理的实施可以概括为认知、规划、试点、推广和制度化五个步骤^[12]。

(1) 认知是企业实施知识管理的第一步,主要任务是统一企业对知识管理的认知,梳理知识管理对企业的意义,评估企业的知识管理现状,帮助企业认识是否需要知识管理,并确定知识管理实施的正确方向。其主要工作包括:全面完整地认识知识管理,对企业中高层管理人员进行知识管理认知培训,特别是让企业高层人员认识到知识管理的作用;利用知识管理成熟度模型等评价工具多方位评估企业知识管理现状及调研分析企业管理的主要问题;评估知识管理为企业带来的长、短期效果;为是否推进知识管理实践提供决策支持;制定知识管理战略和推进方向等。该阶段是企业接触知识管理的第一步,因此需要特别注意两点:① 一方面企业文化和管理模式对知识管理采用何种实施方法有着决定性的作用,因此应特别注意不要忽略企业文化和管理现状;另一方面,由于知识管理的推广需要企业流程、组织和绩效等管理机制的配合,同时也需要深入企业业务层,因而必须得到高层重视,并将知识管理提升到战略高度,才能保证知识管理在企业中顺利推进。② 由于知识管理需要长期的推进,需要对知识管理的效益进行准确量化评估,才能转化为长期发展的动力。因此认知阶段多数企业会邀请外部的一些培训、咨询公司参与,这里关键在于了解业界标杆企业的做法和选择适合自己现状的解决方案。

(2) 在充分认知企业实施知识管理需求的基础上,详细规划也是确保知识管理实施效果的重要环节。这个环节主要是通过通过对知识管理现状、知识类型的详细分析,结合业务流程等多角度进行知识管理规划。在规划中,切记知识管理只是过程,不能为了知识管理而进行知识管理,只有把知识管理充分融入企业管理之中,才能充分发挥知识管理的实施效果。这一过程主要工作包括:从战略、业务流程及岗位等方面来进行知识管理规划;对企业管理现状与知识管理发展的真实性进行分析;制订知识管理相关战略目标和实施策略,并对流程进行合理化改造;进行知识管理系统的需求分析及规划;在企业全面建立知识管理的理论基础。规划阶段的难点主要包括:知识管理和企业战略目标与流程的结合;知识管理与其他管理制度的结合及管理思想的转变;以知识管理思想为基础的业务流程的改造;知识管理的文化氛围的建立;知识管理规划与企业实际情况结合,建立适合企业自身特点的实践形式。建议在规划阶段,由咨询公司和企业高层统一认识,共同参与规划,以更好地确定知识管理实施的解决方案。

(3) 试点阶段是规划阶段的延续和实践,该阶段可按照规划选取适当的部门和流程,依照规划基础进行知识管理实践,并从短期效果来评估知识管理规划,同时结合试点中出现的修正。试点阶段主要工作内容包括:① 进行知识管理模式分析。因每个企业都有不同的业务体系,如生产、研发和销售等,各个不同业务体系的任务特性均不相同,其完成任务所需要的知识亦有不同,因此需要根据不同业务体系的任务特性和知识应用特点,拟订最合适、成本最低的知识管理方法,即进行知识管理模式分析。② 进

行知识体系梳理,对其结果进行分析。考虑到一种业务体系下有多方面的知识,如何识别关键知识,并判断关键知识的现状,进而在知识管理模式的指导下采取有针对性的提升行为,进行知识管理策略规划是一个重要的问题。所以,此阶段的重点是结合企业业务模式进行知识体系梳理,并对知识梳理结果进行分析,以确定知识管理具体策略和提升行为。③ 选择合适的知识管理系统。试点阶段是知识管理从战略规划到落地实施的阶段,根据对企业试点部门的知识管理现状、需求和提升计划的分析,应该考虑引入支撑知识管理落地实施的知识管理系统。根据前几个阶段的规划和分析,选择适合企业现状的 IT 技术,如具有知识管理功能的办公协同系统、知识门户和知识管理系统等。可以说,本阶段在知识管理系统实施中难度最大,需要建立强有力的项目保障措施,做好业务部门、咨询公司和系统开发商等多方面的协调工作。

(4) 在试点阶段不断修正知识管理规划的基础上,知识管理将大规模在企业推广,以全面实现其价值。推广阶段的工作内容包括:通过知识管理试点部门的实践,在企业中其他部门的复制;将知识管理全面融入企业业务流程和价值链;初步建立知识管理制度;全面运行知识管理系统;全面运行学习型组织、头脑风暴等知识管理提升计划,并将其制度化。推广阶段的难点在于对全面推广造成的混乱进行控制和对知识管理实施全局的把握;将知识管理融入业务流程和日常工作;文化、管理、技术的协调发展;知识管理对战略目标的支持;对诸如思想观念转变等人为因素的控制以及利益再分配;建立知识管理的有效激励机制和绩效体系。

(5) 制度化阶段既是知识管理项目实施的结束,又是企业知识管理的一个新的开端,同时也是企业一个自我完善的过程。要完成这一阶段,企业必须重新定义战略,并进行组织构架及业务流程的重组,准确评估知识管理在企业中实现的价值。这一阶段,企业开始意识到知识管理是企业运作的一种战略,而且有必要成为企业综合运作机制的一部分,从而把知识管理全面融入企业战略、流程、组织和绩效等管理体系。在此基础上,知识管理将逐渐演变成企业核心竞争力的一部分,从而有力促进企业每一位员工的发展。

纵观国外知识管理的发展轨迹,结合国内知识管理的应用现状,可以预见在不久的将来,知识管理将逐渐成长成为一种管理思想,进而形成管理变革,如同流程管理一样,将成为体现组织核心能力的关键要素。因此,企业成功实施知识管理对企业核心竞争的增强和企业的长久发展具有重大意义。然而,知识管理从知到行,决不是简单的、盲目的,而是需要涉及多个层面的综合解决方案,企业在推进知识管理过程中,只有透查现状、明确问题,才能合理设计实施路径,发挥出知识管理的真正价值。

11.3.3 企业知识管理系统

在企业中推行知识管理,不但需要员工们懂得知识管理的理念,而且要通过知识管

理系统来使之变为现实。如何实现企业的知识管理系统, 往往是企业知识主管和高层领导所关心的重要问题。近年来, 由于信息技术的进步和 IT 产品的大量出现, 许多 IT 厂商都推出了自己的知识管理系统开发平台, 不少人认为知识管理系统就是一个现成的可以买来使用的软件包。其实, 目前这些 IT 厂商的知识管理开发平台都仅仅是提供了一个可以进一步开发的工具, 要实现企业的知识管理系统, 除了要掌握这些工具之外, 还要对自己将要设计的系统具有清晰的认识。同时, 还必须懂得通过什么样的途径, 才能够减少失败的概率, 有效地完成所规划的知识管理系统。

企业知识管理系统是一个整合的、具有多种功能的、根据企业的业务性质和知识特点所开发的信息系统。就像企业的 ERP 一样, 在开发和引进这个系统之前, 企业的知识总管应当对企业的知识资源以及知识性工作进行分析 and 整理, 并对企业知识系统进行整体性规划。

在系统的构成方面, 面对越来越多的技术可能性应当慎重考虑。软件模块的功能并非越多越好, 应当经过仔细的考虑和筛选, 使得系统具有实用性、友好性和可拓展性三个特性。所谓实用性, 就是要保证员工能够从使用该知识系统获益, 而不是给企业提供一个中看不中用的“花瓶”。所谓友好性, 就是系统应该容易使用, 容易学习, 用户对系统学习没有畏难情绪, 这是系统能够迅速推广的一个重要条件。所谓可拓展性, 就是系统要能够适应企业业务的变化。知识管理系统往往是随着企业需求的增加而扩大的。它需要能随着它所支持的业务一起成长。在系统设计中, 往往从一个核心的需求开始, 逐步扩大其规模和功能。

目前比较通用的企业知识管理系统有 IBM 和微软两套体系。今后还会有更多的体系。作为例子, 下面对这两个体系简单地介绍^[13]。

IBM 体系是以该公司的 Lotus Domino 为核心, 通过 K-Station 等扩展而形成的。它的总体思想是以网络和信息技术为基础, 对组织的知识资源进行系统化管理, 并确定具有专门知识的人员, 建立团队协作的专家网络, 使组织内部所有的人员都能迅速方便地得到所需要的信息与知识。这一体系兼顾了知识活动的三种要素: 人员, 场所 (协作空间), 事件 (包括内容和过程), 并且把积累、共享、交流作为知识管理的主要原则。IBM 体系包括知识传递、协同工作、知识发现与知识路线图、专家认定和专家网络组建、经营智能等知识管理功能。IBM 体系框架结构如图 11.3 所示。

其中的企业知识门户是一个以浏览器为接口的单一访问入口, 通过它来获取、传递和共享知识。它与别的门户不同的地方是具有团队协作功能, 可以使同一组织的人组成团队进行知识活动。此外还可以提供与所得到的信息有关的语境信息。这样就可以实现上面所说的人员、场所、事件的结合。企业知识门户提供了个性化服务, 每个用户都可以自主维护其中自己的主页, 按工作要求组织自己的工作环境。除了日常使用的像电子

邮件、地址簿、日程表等之外，还可以按用户需要加挂新的应用模块。

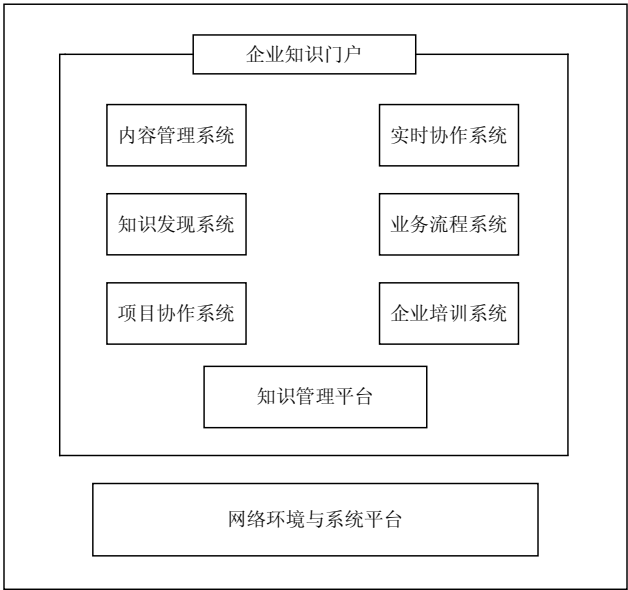


图 11.3 IBM 体系框架结构

这个体系中的内容管理系统能够有效地规划和组织文档，使文档的获取、归类、查找和提取都更加容易，特别是为分布式企业提供了解决方法。它没有像传统的关系型数据库应用时那么复杂。知识发现系统可以自动进行知识分类，并能对文档的价值与用户的切合程度加以确定。它的特点是为人员、场所、事件之间建立语境关系。业务流程系统使企业能够将人工活动标准化和流程化，并能对进度进行跟踪。利用这一系统可以使基于 workflow 应用的开发工作进行得迅速、简单和灵活，并易于维护。在知识沟通、交流方面，实时协作系统和项目协作系统提供了即时通信服务、会议服务和集成服务的功能，具有很好的可伸缩性、可管理性和安全性。它还有一个工具（称为 QuickPlace）可以用来针对工程项目或时间，迅速建立网站，将协作交流、文档共享、用户管理、日程安排、任务分派等集成一体。企业培训系统可以支持网络学习、移动学习和实时培训。

这一体系在进一步发展时，应尽量发挥其协作支持能力，并与以分析为主的经营智能相结合。

另一个体系是微软公司开发的。它的思路是通过组织、过程、技术三方面来提升组织的核心竞争力。从过程上来说，要保证知识管理与经营过程相匹配；从组织上来说，要克服知识共享的障碍，培育创新精神；从技术方面来说，要使用户使用它所熟悉的工具进行知识共享活动。微软提供的知识系统体系结构框架如图 11.4 所示。

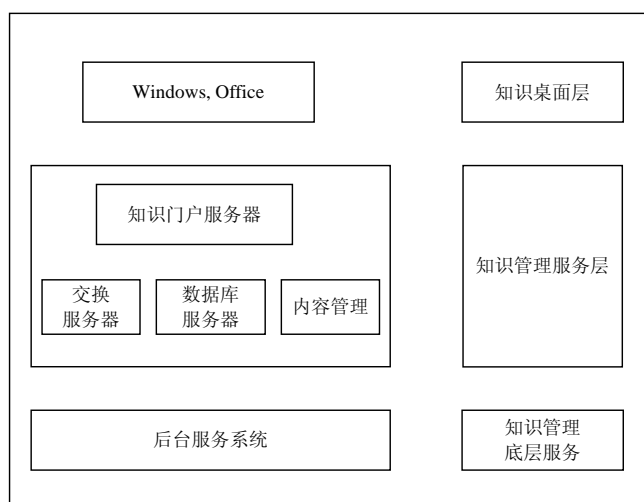


图 11.4 微软知识系统体系结构框架

从图 11.4 中可以看出，它是一个三层结构。最上层是知识桌面，它是以操作系统（Windows）与个人工作效率软件（Office）为基础的。这是进入系统的一个门户。它的特点是采用了用户所熟悉的接口和操作方法，来运作企业的各类知识资源。

在知识管理服务层，含有下面四种功能：

（1）内容管理。这部分管理功能可以对存储在知识库中的知识内容进行采集、整理与组织，并与其他部分如知识桌面、交换服务等相结合，又可以进行分类、发布和文档管理。此外，它还支持版本管理、文档审批、邮件路由等功能。

（2）协同工作。这一部分管理功能以交换服务器（微软的 Exchange Server）为核心，与知识桌面层的 Office 相结合，提供了协同工作的环境。它能够让用户使用他们最为熟悉的工具，进行合作，完成知识应用和知识创新工作。其中包含了日程表、工作安排、在线讨论等环节，以及用网页文件夹来帮助团队进行合作的工具。此外还可以跟踪 workflow。

（3）经营智能。这部分管理功能是以 Office 为接口，而以数据库服务器（微软的 SQL Server）中的分析模块为基础，能够使各层次的知识工作者进行在线分析处理（OLAP），以及其他分析研究。

（4）知识门户服务。这是微软知识管理体系中的核心环节。微软的 SharePoint Portal Server 就是这种服务器。它把文档管理和搜索功能与日常使用的工具集成在一起，使得 Office 与万维网浏览器乃至 Windows 中的资源管理器集成在一起，使得用户不必在搜索、组织和管理信息的工作上花费更多的时间。

文档管理需要一个合理的框架，来对大量的信息源进行组织和应用。知识门户服务器就能提供这样一个架构，通过它可以进行文档配置、文档发布、审批流程和网上讨论等。

在索引和搜索服务方面,服务器提供了全文检索、按主题或类别进行浏览、文档分类、订阅以及进行最佳匹配功能。它允许同时在多个信息源中进行搜索。

在知识管理底层服务中,则包含各种后台服务系统。

微软体系中还提供了一种可由用户自己创建保存所有项目信息(包括文档、联系人、任务、讨论以及其他信息)的中央数据库的服务,称为 SharePoint Team Services。它使任何类型的工作组都能使用网络分享信息。利用它可以即刻组建团队,使用全功能的网络站点,还可以进行团队互助和团队管理。

上面简要介绍了企业知识管理系统的两个体系,其目的在于通过它们可以了解到集成的知识管理系统的构架和组成要素。现在还有许多系统也或多或少具有多种功能,这里不再介绍。

11.4 政府知识管理

知识管理的概念源于企业管理,并且成为企业管理中的研究热点。知识管理对企业生存与发展的重要性,引发了人们对政府知识管理的关注和思考。企业知识管理理论对推行政府知识管理具有重要的借鉴意义,是政府知识管理发展的基石。

11.4.1 政府知识管理的含义

政府知识管理是指利用现代信息技术,充分发挥政府知识资源的作用,实现知识的获取、共享和应用,提高行政人员和行政组织的学习和创新能力,提高政务工作效率,增强管理、服务和领导决策能力,促进国家、社会的繁荣和进步的一种管理模式。

政府中对知识的管理主要包含对知识获取的管理、对知识共享的管理和对知识应用的管理三个方面^[14]。

(1) 对知识获取的管理是政府实现知识管理的第一步。在政府知识管理中,获取知识的途径有内部渠道和外部渠道两种。所谓内部渠道就是指通过政府机关内部的工作人员通过交流和知识挖掘得到的新知识;外部渠道是指从政府机关外部获得的知识,主要包括专门的政府智囊团、科研机构、各类咨询公司以及公众。

政府对知识获取的管理主要表现在三个方面:① 加强政府部门内部的交流,通过会议、讲座等形式获得内部公务员掌握的经验知识;② 充分利用政府智囊团、科研机构和咨询公司等外部资源,通过合理的合作方式,有效地获取他们掌握的知识;③ 通过定量或定性的信息和知识处理方法,从已经掌握的丰富的政务信息资源中挖掘出有用的知识,这种方式应该成为政府获取新知识的主要途径。对于知识的这种挖掘,需要有相关的基础数据库的支撑,因为只有对大量的电子化的信息才能利用计算机进行有效处理,才能

从中挖掘出有用的知识；对于大量的存储于纸上的信息，目前还很难应用信息技术来获取知识。

(2) 政府内部的知识共享可以分为两个层次。第一个层次是政府内部不同职能部门之间的“横向”知识共享；第二个层次是相同政府职能部门的不同等级之间的“纵向”知识共享。通过不同职能部门之间的“横向”知识共享，可以有效弥补部门之间的信息和知识鸿沟，有利于从更全面、更高的角度来看待和解决遇到的社会和经济等方面的问题，促进整个社会的可持续发展。而通过政府职能部门之间的“纵向”知识共享，可以使得部门内部有效的经验知识得到广泛传播，为整个部门所共有，推动整个部门的发展。

政府对知识共享的管理包含两个方面的内容：① 建立一个统一的知识共享硬件平台，使得政府的不同职能部门之间或相同部门的上下级之间的知识共享具有实现的基础；② 营造知识共享的“软”环境，通过有关法律、法规和激励政策的制定，指导、鼓励政府的知识共享。

(3) 知识的获取和共享都是为知识的应用做准备的，政府中对知识的应用主要是为了实现政府的管理和服务这两种职能。政府管理职能的实现离不开各种政策法规的制定，离不开各种决策行为。在政府中，知识应用的主要目的之一就是更好地为领导的决策提供支持功能，使得决策更科学化。在知识经济时代，政府对企业和公众的服务职能日益凸显，社会公众对政府部门的办事效率、服务水平也提出了越来越高的要求。如何满足社会公众对政府服务职能的需求，提高政务事务的处理效率，提供满足公众需求的服务项目，是当前政府面临的问题和挑战。这些问题的解决需要政府有针对性地获取有用的知识，并合理地应用这些知识。离开了知识的应用，政府的服务职能也难以很好的实现。

政府对知识应用的管理主要表现在三个方面：① 根据用途对知识进行分类管理。在应用知识之前，必须首先识别知识的类别和作用；② 建立一个应用知识进行科学决策的平台，鼓励各级领导在进行决策时应用各种相关知识，包括领导自己的经验知识，也包括通过知识获取手段得到的其他知识；③ 对各职能部门在知识获取、共享和应用方面的管理情况进行评估，并作为衡量其成绩的指标之一。

11.4.2 政府知识管理的实施

知识管理对政府职能的转变提供技术支持和方法论基础，政府知识管理的目标是：充分利用现代信息技术，构建先进的电子政务系统，挖掘政务信息资源中隐含的知识，创建良好的政府文化和组织形式，引导知识创新，实现知识共享，并通过对共享知识进行有效应用，提高政府工作效率，提高决策的科学性，提高政府对公众的服务能力和公众对政府的满意度。

政府知识管理实施的条件包括以下三个方面^[15]：一是要提供制度保障。在政府部门

推行知识管理,首先要提高政府官员(特别是主管领导)对知识管理的认识,明确知识管理的重要地位。政府财政部门应加大对知识管理项目的资金投入,确保政府对知识管理和信息资源共享的措施到位,推进知识管理的实施。二是要搞好信息基础设施建设。在政府部门推行知识管理要求有关机构和部门在对政府信息化作出科学合理的发展规划基础上,加强政府知识管理平台(信息基础设施)建设,从而建立一个高效的知识管理系统。三是要搞好人才保障。在政府部门推行知识管理,本质上是知识与信息技术契合的过程,公务员在知识管理系统中,应该有较全面的信息管理能力、信息技术的使用能力。因此,必须不断加强对知识管理人才的培养。

政府知识管理是从企业管理中引入公共部门的一种新兴的管理思想和理念,目前政府知识管理的运作体系尚未形成统一的框架,其具体的实施过程需根据不同政府部门特点及自身资源条件的不同而有所差异。结合国内外政府知识管理的成功实践,在政府部门建立完善的知识管理体系具体来说可从以下五个方面进行^[15]。

(1) 构建扁平化的学习型政府组织。根据知识管理的特点,政府部门需要调整组织结构,使组织结构与学习型政府的特点相适应。首先,政府需要设定主管知识管理系统的领导,负责创造一种能够促进学习、积累知识和信息共享的环境;监督并保证知识库的内容质量和正常运行;加速知识继承,催化新知识的产生;进行创新和发挥集体的创造力。其次,要构建扁平化的学习型组织,使组织更适于学习和建立开创性思考方式,使之有利于公务员发挥主动性,缩短知识传递的时间和空间。通过开放型、学习型、成长型的网络组织不断交流知识,政府内部知识共享机制会从中产生。

(2) 构建知识共享的激励制度。在政府部门推行知识管理时应该做到机制先行,以使知识管理系统建立后的运行有制度保障。政府知识管理必须建立知识的生产、传播、共享和运用的激励机制。首先,要建立知识运营的激励机制,这主要由知识交流机制、知识管理目标发布制度、创新失败宽容机制构成。其次,是建立知识的绩效机制,需要对有关人员申报的知识管理成果进行审查和评定。这其中包括知识成果审计制度、计算机综合评价系统等。再次,是知识奖惩机制。它包括知识薪酬支付制度、人员晋升制度、知识署名制度、知识培训制度和知识老化型职员淘汰制度等。

(3) 构建知识交流的信息网络系统。该系统包括政府外部网络系统和政府内部网络系统。构建知识交流信息网络系统时,政府可以建立临时任务小组、课题组和矩阵结构小组促进知识管理。在这些非正式的职能小组中,知识管理信息网络系统中的各元素可以是研究所、大学、实验室、企业、课题组和个人,这种形式可以使他们在组织中加强交流,也可使组织的灵活性与适应性能很好地结合起来。把非正式的职能小组用信息技术网络连接起来,每个元素都是网络中的一个信息点,其中最为活跃、创新能力最强的一个或几个节点就成为知识管理创新节点。同样,政府管理创新目标最有可能在创新节

点上实现。为了解决公务员能够“在最恰当的时间找到最恰当的知识”这一问题，最好的办法就是建立有序的、便于查找的知识库，编制知识地图，然后再通过知识索引，让公务员在知识库中找到“最恰当的知识”。这个系统就是政府内部网络系统。这个内部网络系统主要包括硬件和软件要素、知识编码和知识库。硬件和软件主要包括计算机、网络、数据库和相关技术。构建这样的—个内部网络系统，使公务员易于进入知识库也便于查询，获得精确、及时、可靠的知识和信息。

(4) 构建知识积累创新的管理系统。知识管理系统是一个能够不断学习、创新的循环系统。在这个知识循环系统中，政府有关部门和人员首先利用知识发现方法从已有的信息和知识中发现新知识并进行采集；知识表示系统将新知识以用户容易理解的方式表达出来，得到用户的确认后存储到知识存储系统中；存储后的知识通过网络和其他途径实现知识共享；通过知识共享机制和实施，有关部门就可以运用知识库的知识；通过运用知识又可以实现不断创新，在下一个循环中，知识发现系统可以利用上个循环中新发现的知识进行新的知识发现，从而实现了知识管理系统的不断学习和创新功能。因此，知识管理系统运行的状况直接决定了政府的知识学习能力、创新能力和决策能力。

(5) 营造促进知识管理的行政文化。上面提到，政府部门进行知识管理时，必须充分发挥集体智慧，重点是充分发挥公务员个体的聪明才智，实现知识创新。因此，必须在组织中注入个人价值与自我责任感，实现个人价值与群体创造知识形式的有机协调，创造尊重人，有利于个人价值的实现，保持群体间互相协作，共享信息和知识，减少并消除内耗的环境，使政府公务员的创造力能够最大限度地发挥出来，并形成一种集体的创造力和创新能力，进而更好地发挥自己的积极性和创新能力，实现自己的人生价值。

总的来说，在政府部门推行知识管理的目的在于促进知识共享和知识应用。知识共享能够促进知识的创新和传播，实现知识的不断增值，并在日益加剧的、不连续的环境变化中增强政府的管理能力和服务能力。



思考与讨论题

(1) 什么是个人知识管理？如何实施个人知识管理？常用的个人知识管理工具有哪些？

(2) 项目管理知识体系中包括哪些知识？如何对项目各阶段的知识进行管理？在项目管理中，专家知识应该如何获取？

(3) 如何实施企业知识管理？调研一个实际运行的企业知识管理系统，描述系统的体系结构。

(4) 知识管理在政府管理中有何重要作用？如何实施政府知识管理？



参考文献

- [1] 储节旺, 周绍森, 谢阳群, 等. 知识管理概论[M]. 北京: 清华大学出版社; 北京交通大学出版社, 2006.
- [2] 邱均平, 等. 知识管理学[M]. 北京: 科学技术文献出版社, 2006.
- [3] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [4] 王众托. 知识管理[M]. 北京: 科学出版社, 2009.
- [5] Dorsey P. Personal knowledge management: educational framework for global business[EB\OL]. http://www.millikin.edu/pkm/pkm_istanbul.html, Tabor School of Business, Millikin University, 2001.
- [6] Gery G. Electronic performance support systems[M]. Boston: Weigarten Publications, 1991.
- [7] 王众托. 项目管理中的知识管理问题[J]. 土木工程学报, 2003, 36(3): 1-6.
- [8] 中国(双法)项目管理研究委员会. 中国项目管理知识体系[M]. 北京: 电子工业出版社, 2006.
- [9] Mikolajuk Z. Knowledge management in development projects[EB\OL]. <http://www.bellanet.org/km/attach/km4dev-1/42634/KMdevprj.doc>.
- [10] 仇元福, 潘旭伟, 顾新建. 项目管理中的知识集成方法和系统[J]. 科学学与科学技术管理, 2002, (8): 36-39.
- [11] 江文年, 杨建梅. 企业知识管理方法论研究[M]. 科学出版社, 2006.
- [12] MBA 智库百科. 企业知识管理[EB\OL]. <http://wiki.mbalib.com/wiki>.
- [13] 王德禄. 知识管理的 IT 实现[M]. 北京: 电子工业出版社, 2003.
- [14] 李英, 蒋录全. 电子政务与政府的知识管理[J]. 华东理工大学学报(社会科学版), 2005, (1): 125-128.
- [15] 王志凌, 谢宝剑, 邓晓海. 基于知识管理视角的政府管理创新[J]. 中国行政管理, 2007, (4): 28-30.

第 12 章

知识系统的开发

内容提要

本章讲述的是知识管理系统开发的战略考虑和实施步骤。首先从知识管理的战略出发，确定系统的愿景和目标，然后研究开发的步骤。其中，对信息基础设施的组成也进行了介绍，还介绍了知识流程外包的特点和内容。最后介绍了一类隐性的知识管理系统——学习型组织。

本章重点

- 知识战略
- 知识管理系统的开发步骤
- 知识管理系统的信息基础设施
- 一类隐性的知识管理系统——学习型组织



12.1 知识项目与知识系统

在一个企业或者其他组织中,由于知识已经在不同的部门和不同的团组中有所应用,有的还有所创新,一些知识密集型的组织对知识的应用和创新比较自觉(当然可能只着重于科学技术知识),人们已经感觉到知识活动的存在。实际上这时已经自觉不自觉地形成了知识系统。在当前这样一个科学技术的作用日益显著、竞争日益激烈的形势下,企业都想通过知识的运用和创新提高自己的核心竞争力,因此总会想到如何开发智力资源,建立知识经营和管理项目。这些项目有一定的针对性,如开发新产品,做出投资的重大决策,等等。每一个项目都是一次性的,但都会不断增强和改进企业的知识系统。

我们这里所谓的知识系统,不是仅指建立在信息技术基础上的以管理显性(可编码)知识为主的知识管理系统(这是西方所谓的第一代知识管理的主要形式),而且是指以人为中心、同时开发和运用显性和隐性知识的系统,这个系统不是人围着计算机转,而是计算机等信息工具随着人转。系统的组成不仅包括有形的知识和工具,还包括无形的知识资产和最重要的知识源泉——作为知识创造者的人。

知识系统的任务一般包括:

- (1) 获取和生成知识;
- (2) 对知识编码和标准化;
- (3) 进行知识组织与处理;
- (4) 支持新知识的生成;
- (5) 进行知识转移和传播;
- (6) 支持组织学习。

在本书的第3章曾经提到,广义的知识系统实际上包含知识处理和知识管理两个层次,或者说,从功能着眼,知识系统包含知识处理和知识管理两个子系统。而知识管理层又分成如下三个层次:

- (1) 战略管理;
- (2) 战术管理;
- (3) 运作管理。

这与一般工业生产的管理是类似的,而且各层也与实物生产管理的各有关层次有一定的联系。特别是知识管理系统中的运作管理层和知识运作层是难以分开的。例如,在产品设计工作中,计算机辅助设计(CAD)系统是知识处理系统,而产品数据管理(PDM)或产品生命周期管理(PLM)系统是知识管理系统,两者之间联系紧密,两类系统是建立在一个集成环境之中的,难以分开。

如果从更高的层次来看,信息管理系统的任务是对信息进行采集、处理、存储、传递和利用,而知识管理系统则是对知识进行采集、处理、存储、传递和利用。由于信息是知识的载体,所以知识管理各环节与信息管理的各环节的联系是紧密的,有些就是融为一体的。从信息管理的角度来看,知识管理可以认为是信息管理的延伸和扩展;而从知识管理的角度来看,信息管理是知识管理的前奏。信息管理的某些内容是表层的,而知识管理的内容是深层的。

知识经营管理项目的建立和知识系统的开发,就是在组织现有的条件下,按照知识战略,有步骤地实现这些措施。

知识项目或知识系统的开发由于任务和要求不同,既可以是整个组织从长远着眼,建立一个实施整体知识战略的大型项目或系统,也可以是为了某项具体任务开发的一个局部项目和系统。前者与知识战略联系紧密,而后者是在知识战略的指引下更多地从战术要求出发。前者着眼于长久的经常性的工作,后则更关注于具体运作。

现在不是每一个组织的领导或主管人员都认识到什么是知识管理以及它的作用,一个需要开发的知识项目可能是从某一项业务需求(如新产品开发、进行重大投资决策)出发,这时也可先不急于把它称作知识项目,还是叫做新产品开发项目、重大决策项目,而在实际工作中,针对具体要求将知识活动贯穿在其中,等到项目有了成果,人们开始从实践中认识到知识的作用了,再冠以知识之名。这样可以避免由于当事人对知识管理不理解而遇到不必要的阻力。

12.2 知识系统开发的战略考虑

知识项目和系统的开发首先要有战略上的考虑。

12.2.1 知识战略与组织总战略

在知识经济时代,知识的运用和创新必须提到战略的高度来考虑,也就是说组织要有知识战略,而知识战略又是与总的战略分不开的。组织(如企业)总的战略是为了实现组织(企业)的总体目标而对所有的资源进行合理的配置和采取的行动总计划。

研究知识战略的前提是:当前知识作为一种重要资源,需要组织在制定战略时认识到它的重要作用,采取相应的战略措施,合理配置资源,特别是知识资源,以提升本组织的核心竞争力和对市场变化的快速和灵活的反应能力。

核心竞争力指的是企业所独具而在一系列产品和服务中所必须依赖的关键性能力。核心竞争力通常是企业多年在一个领域中从事工作与管理形成的,代表了企业的长期积累和组织特点。这种能力不是拥有的单项技术或技能,而是综合了一组技术和技能。从

知识的视角来看,核心竞争力是组织中累积性的知识和经验,特别是如何协调不同的生产技能和有机结合多种技术流派的知识和经验。

核心竞争力包括下面这些能力:产业洞察力或预见能力,也就是能把市场机会与现有资源联系在一起的非凡能力;系统组织能力,即确定企业从事的核心活动范围的能力;运作能力,也就是在经营中怎样实施的能力。这些能力中洞察力是关键性的。

缺乏核心竞争力的企业也许一时还能生存下去,但难以发展,不能在市场上占领先地位。有了核心竞争力,企业就能在环境的迅速变化与市场的剧烈竞争中立于不败之地。由于企业的核心能力具备鲜明的个性特征,所生产的商品或提供的服务能使顾客得到额外的满足,所以能在市场中独树一帜。

核心竞争力还能在未来的发展中创造出新的产品或系列服务,打开新的市场。

与企业的核心竞争力相近的另一个概念,是企业的战略素质。战略素质是一个企业能将对竞争至关重要的价值流经常置于优异水平之上的素质。某一种核心能力表现为价值流上某一点的技术和技能,而战略素质则着眼于整个价值流。核心竞争力和战略素质构成了企业战略思想的中心。

上面说的是企业的知识战略。在政府、教育、文化领域的知识管理中,同样也有知识战略问题,同样也要服从组织的总体战略。例如,政府的知识战略需要立足于服务功能的优化和政府职能的转变;教育的知识战略要立足于全面贯彻教育方针和培养师生的创造力;图书馆的知识战略要立足于优化图书馆的业务工作和管理活动,以更好地满足现代社会的需求。

知识战略正是在这样一种全新的战略思想指导下,为实现组织的总体战略,制定知识管理的愿景、使命、目标和战略措施,将组织中与知识有关的活动如信息资源管理、技术管理、人力资源管理等集成起来,进行综合的经营和管理。知识战略是组织总体战略的一个组成部分,或者说是总体战略中一项专门性的战略。

一个组织的知识战略,又是与国家的知识战略和知识政策密切相关的。国家的知识战略是指在国家层面上制定的怎样依靠知识的有效运用和创新来提升国家实力的战略规划,知识政策则是围绕着知识战略制定的相关政策,具有导向作用。国家的知识战略包括知识资源战略、知识人才战略、知识环境战略、知识法律战略等。企业等组织的战略必须在大方向上与国家战略相一致,并且符合知识政策。

12.2.2 知识战略的组成

组织(企业)的知识战略应该包括下列六部分内容:

- (1) 知识经营管理的愿景、使命、目标;
- (2) 对知识经营管理的认识原则;

- (3) 发展的机遇与面临的挑战, 以及应对措施;
- (4) 所需资源及其配置;
- (5) 政策措施与计划;
- (6) 效益分析。

一个好的知识战略应该具备下面三个特点:

(1) 前瞻性。由于人们对知识的重要性的认识是逐步加深的, 特别是目前, 中国处在工业化进程的中期, 对先进技术与先进经营管理思想的认识和采用要有一个接受的过程, 但作为高层领导要有一定的预见性和前瞻性, 才能做到与时俱进。

(2) 可行性。必须考虑组织的实际情况, 做到预定的目标经过努力是能够达到的。

(3) 灵活性。在内外环境发生变化时, 能够及时加以调整, 以适应外界需求。

针对企业的不同情况可以采用不同的知识战略。美国的生产力与质量管理中心(APQC)曾经将知识管理战略分成下列六类。

(1) 以知识经营管理做为企业战略。这种战略着眼于知识对企业成长、竞争力提高的关键作用, 强调将知识的获取、创新、组织、共享应用到企业所有的业务和管理领域, 使知识得到最有效地利用。推行这种战略时, 组织坚决相信知识的经营管理会对本组织的生存和发展起到决定性的影响, 因而不断地把知识结合到产品、服务和业务流程中去。

(2) 知识与技能最佳传递性战略。这种战略强调把知识传递到能够产生最有效行为的地方, 强调以团队、网络和各种联络关系为手段来传递知识, 以改进产品和服务。这是当前最常见、最普遍的一种战略。

(3) 以用户为中心的战略。这种战略面向用户、针对用户的特殊要求不断地设计、开发和调整自己的产品或者服务, 协助用户取得成功。其做法是获取、开发和传递有关用户需求、偏好和业务方面的知识, 从用户那里学习, 理解用户的需求, 以更好地满足用户的需求。

(4) 个人对知识的责任型战略。这种战略强调员工个人对于自己责任范围内与知识有关的投资、创新、有效利用以及竞争等情况负有责任, 要求员工把最佳的知识用于业务工作。这种战略相信员工能够通过“拉”的技术主动获取知识而不需要组织“推”给他们。员工自己会不断充实自己, 使自己的知识能够胜任具有高度挑战性的工作。组织只是培育一种有利于知识利用和创新的文化氛围, 并有健全的激励制度。

(5) 知识资产管理型战略。这种战略强调对知识资产进行统一的集中管理, 充分开发和利用组织的专利、技术、操作和管理技能及客户关系等知识资产, 提高其使用的有效性, 并进行知识营销。

(6) 知识创新型战略。因为这种战略相信创新是增长的核心, 创新能提高组织的竞争力, 所以这种战略强调组织学习, 鼓励员工的知识创新活动。通过基础与应用研究以

及开发活动来创造新知识,要求员工不断获取新的更先进的知识。

上面几种战略是当前一些组织已经采取并行之有效的战略,至于一个企业在选择自己的战略时究竟采用哪一种,则要看组织的性质、知识工作的类型和环境条件来确定。也可以同时采取上述两种方案进行综合集成,或者创造出其他新方案。

在战略选择上,还有一个问题要考虑,就是以信息技术支撑为重点还是以对人的管理为重点的选择。前者有利于显性知识的充分共享,而后者有利于隐性知识的充分发挥作用;前者知识流失的风险小而后者风险大;前者在信息技术设施上投资大而后者投资小。当然最好是两者兼顾,因为这两者是无法分开的,不能想像在开发和运用信息技术时离得开人员的隐性知识作用,而在发挥人的知识作用时离得开一定的技术支持。但总的来说还是以一个为主,具体以什么为主需要按照对主要知识类型的需求来确定。

12.2.3 知识战略的实施

实施组织知识战略的重要举措,包括如下五部分内容。

(1) 构建组织架构。由于知识管理的责任是分散的,所以需要建立一套有效的组织体系,来支持企业实施知识管理。这一体系中一定要有专门的领导人负责知识管理活动,而不是由专管其他业务的领导来兼管一下。负责的领导要制定知识管理的规划和计划,协调各种知识活动。与知识有关的任务,如技术改造和产品创新等,要有专门小组负责。

(2) 构建和运行组织的知识储备设施——组织的知识库。正如前面所说,这里所说的组织的知识库是一个抽象的概念,它不但包括像计算机数据库、文件夹这样一些具体的东西,还包括存放隐性知识的个人头脑与集体的默契配合。对于数据库和知识库,要有一定的资金投入,这不仅包括创建阶段的成本,还包括常常为人所忽视的数据与信息的采集成本。

(3) 构建与运行知识网络。这里所说的知识网络不但包括具体的信息网络,还包括像专家网这样的社会网络。对于与知识网络有关的基础设施,如信息网络和专家网络等,都要设立专门的组织来管理。

(4) 对企业(组织)的知识进行经营,使其发挥作用,创造价值。这包括对知识产权的保护,知识投入产出的分析与知识过程的优化,以及知识的转让等。尽管对知识管理的效果难以定量评估,但还是要建立知识管理的评估系统,应该研究建立面向知识资产的评估体系,并随时加以评估和检测。

(5) 营造一个有利于知识创新的文化氛围,培养员工的良好职业道德、团队精神和集体荣誉感,还包括建立激励机制,鼓励合作交流。

在制定组织的知识战略时,应该处理好下面五方面关系。

(1) 知识战略要服从总的战略,而不是脱离总的战略。这在前面已经讲到过。因为

知识工作的最终目的还是为了提高组织的核心竞争力和应变能力,而且知识工作只有融于组织的总任务才能达到目的。如果知识战略自己孤立地另搞一套,就会得不到其他方面的配合与支持,缺少这种支持是无法实现自己的任务的。

(2) 知识战略的着眼点包括知识组织和知识团队,而不仅仅是知识工作者的个人。知识战略措施既要发挥知识工作者个人的作用,更需要发挥知识团队的作用,因为有效的知识应用和创新依赖于组织知识,而组织知识是掌握在集体手中的。

(3) 知识战略既要重视技术的作用,如信息技术网络的作用,更要重视人的作用,如知识工作者之间关系网络的作用。

(4) 知识战略的重点是隐性知识的采掘、转化和应用。

(5) 知识战略必须与原有的企业核算和评估体系有所联系,使得知识应用的成果能够体现在企业的效益计算中。

12.3 知识系统开发的步骤

这里提出一个开发知识系统的一般步骤,可供企业在开发各种规模的系统时参考。下面列举的步骤不一定完全需要按部就班地执行,但是其先后逻辑关系却值得参考。

知识系统开发是在组织的知识战略已经确定的前提下开始的。开发工作可分为下列九步骤。

1) 进一步明确知识战略

这在 12.2 节已经讲过了,此处从略。

2) 对现有知识资产进行分析

在系统开发的一开始,就应该先对组织已有的各种知识要素进行调查了解。按照本章 12.2 节谈到的原则制定组织的知识战略之后,就已经能够对本单位的知识作用和需求有一个概略的了解,这时应该通过知识资产的各个成分来分析组织究竟已经具备了哪些知识能力和知识积累。以此为出发点,就很容易往下进行了。

3) 按照组织的知识战略确定知识工作的目标和任务

知识系统总是要为具体的任务服务的,组织的知识战略确定了一般的目标,还得落到实处。也就是说,应该制订规划,把具体目标和任务加以确定;需要进行优势和劣势、机遇和威胁的分析(SWOT),找出现状与需要之间的差距,研究消除差距的主要措施。在这个阶段,对于进行的调查分析要做出评价,甚至进行知识资产的评估,既要找出具体的资金的缺口,同时也要发现潜在的能力。

4) 研究确定知识系统的各种体系结构

实施一项知识项目其实质就是构建一个知识系统或者改造一个知识系统,为此需要

研究确定系统的各种体系结构。本书第3章曾经讨论过体系结构的作用和内容,有了体系结构,明确了各方面的关系,就容易进行组织、人员、技术等各方面的安排。体系结构的制定要避免两种偏向:一种是勾画得太笼统,使人抓不住要点;另一种是描述得太具体,代替了后来的具体设计。一般说来,能够把各部分的功能和相互关系描述清楚,借助结构图等形式表述出来就可以了。还有一个问题是如何在以信息管理为主线和以人的管理为主线的选择和折中方面,这在一般情况下应该根据需求特点和本身条件加以考虑。我们主张把两种思路集成起来,但在某些局部总要解决以谁为主、以谁为辅的问题。应该加以研究确定。

5) 分析有关知识运作和管理组织,进行组织的设计和人员的安排

对各层次的知识管理组织和人员进行安排配备。组织变更和人员调配是很敏感的问题,因为在知识管理机构 and 知识运作机构之间,知识管理人员和知识工作者之间存在着多方面的关系。表面上是工作分工、协调问题,实质上常常会涉及权力分配和利益分配问题。这些人为因素如果在开始时没有解决好,事后会因此产生许多阻力和障碍,影响知识的运用和创新,所以需要领导亲自参与。这时还需要考虑后面将要讲到的实践社群的形成问题。虽然这种社群是在实践过程中自然形成的,但是预先要有一个考虑,如已有的社群是否会发展成为与知识创新有关的社群,应该为将来的社群准备什么条件,等等。

6) 设计知识流程

研究知识的运作和知识创新的流程,需要按照知识战略或者知识运用与创新的具体项目安排。前面曾经分析过,知识运作和运作管理的过程包括:确定知识需求,进行知识识别,进行知识收集,进行知识选择,进行知识保存,进行知识传播与共享,利用知识,转化知识,创造新知识。这其中的每一项任务或过程都有一定的工作流程。它们中有一些如新知识的生成等是一个复杂系统中的自组织过程,带有混沌特点,很难用有序的流程来描述。但是还有一些过程是可以建立科学化的流程的,应该分析和设计这些流程。这对下一步构建信息基础设施也是有用的。在此基础上还可以制定一些工作导则和规范,确定一些激励机制等。

7) 构建信息技术基础设施

现代化企业等组织的知识系统应该建立在先进的信息技术的基础之上。虽然以人的管理为主线不如以信息管理为主线的模式对信息技术依赖那么多,但是为了加强人际沟通,还是少不了依靠信息技术的帮助。即使对知识工作者个人的工作,也还得依靠信息技术工具支持。所以在开发过程中,不能忽视信息技术系统的建设。

对于信息技术基础设施的构建,应该摒弃过去的一种偏见,以为这只是购置计算机和通信工具,架设信息网络,最多还包括购买或开发软件。其实更困难的还是构建信息系统的工作方式,使其为知识的运作和创新服务。应该从知识存量着眼来研究知识在信

息系统中如何积累, 以及从流量着眼来研究知识在信息系统中怎样流转。

在本书的第 9 章曾研究过知识系统中的信息支持工具问题, 这里就不再赘述了。

8) 知识系统的实施

知识系统的实施包括成立组织, 任命相应的人员, 采购和安装、调试信息技术基础设施。这些有形的任务容易被人们所注意, 但还有一些无形的任务, 如组建专家网络, 建立对外的知识工作关系, 考虑组织文化的形成措施等, 也要及时进行。在实施过程中, 为了取得经验, 可以在一个小范围内针对某一项有关知识运用或知识创新的工作, 进行试点。

9) 进行系统的评估和总结, 明确改进的方向

上面这些步骤, 适合一个企业或其他组织开发总体系统或某一具体系统时参考。但是知识工作不是一劳永逸的, 知识系统也得不断地进化, 一旦建成一个较为完整的系统, 以后的发展大多是渐进的。但上面的一些步骤还是可以参考的。

12.4 知识系统的基础技术设施

从前面所讲到的可以知道, 系统的功能应该包括如下四个方面:

- (1) 提供用户一个单一的入口以便获取知识;
- (2) 对企业所有的知识资源进行保存与管理;
- (3) 对知识流程进行管理;
- (4) 提供知识的协同工作环境。

一般来说, 正如本书第 3 章所指出的, 这类系统分成四个层次:

- (1) 表示层, 这是系统与用户的接口;
- (2) 应用层, 这是对知识应用和知识创新进行管理的实现层次;
- (3) 功能层, 这是应用层与数据层之间的联系层, 主要为应用层提供服务;
- (4) 数据层, 这是用于进行企业各种知识资源的保存和管理。

知识系统的结构如图 12.1 所示。

我们可以自上而下地对各层次的作用和组成要素进行一些分析。

1. 表示层

表示层是用户与系统进行联系的接口。它的功能是响应用户的操作指令, 迅速导向有关的应用模块, 进行处理, 并将处理结果进行显示。它应该是用户进入系统的唯一入口。

简单的表示层就是通用的图形接口, 大型的、复杂的系统则是用企业知识门户来充当入口。它可以使企业的员工、客户和合作伙伴都能够访问、获取、交流与他们有关的知识。

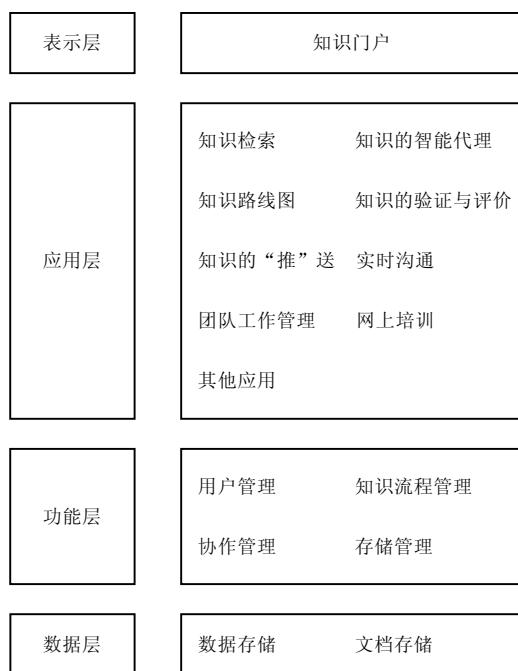


图 12.1 知识系统结构图

对表示层的要求是它应该使用单一的访问方式，一般是使用浏览器进入系统，进行操作。在一次登录进入后，可以访问不同的模块。

它应该能够进行个性化的定制，用户可以自己定义自己的角色，选择相应的页面，按自己的偏好过滤信息。

这个层次不仅是一个入口，而且能够把已有的系统加以集成，特别是知识路线图的集成和团队工作平台的集成。

2. 应用层

应用层直接决定了整个系统的能力。它包括本书第 3 章所说的知识运作和知识管理的各种应用模块。这些模块有的是通用的，在任何企业或组织中都能够使用；有的是专门针对某一企业或者它某一项任务特点的；有的是针对个人使用的；有的是针对团组使用的；还有的是面向组织全体人员的；以及面向企业外的人员如用户、供应商等的。下面列举一些有代表性的模块。

1) 知识检索

与用户最先接触的应该是知识检索模块。它为用户提供了一个统一的检索工具，并能够支持多种检索方式。在今天的技术水平上，全文检索已经没有什么困难，但利用关键词检索可能得到的结果会过于繁杂而缺少针对性，因而出现了许多利用智能技术提高检索效率的方法，以及具有学习功能和在检索过程中不断改进检索方法的工具。

2) 知识的智能代理

智能代理可以认为是一种软件机器人,利用它可在后台运行,接收用户的知识查询请求,代替用户搜索文档。由于它能够收集用户操作记录,对用户的操作规律和知识需求进行分析,因而能更切合用户要求去进行知识收集。

3) 知识导引(知识路线图)

知识路线图是一种向导,能够指引用户找到所需要的知识。它不仅可以指出可编码的知识在什么地方,而且可以指出具有知识的人在哪里,这样就可以找到隐性知识的来源。知识路线图给出的是知识的位置,而不包括知识的内容。所以,它只起到索引的作用。系统中应该具备建立知识路线图的工具,利用它可以进行知识的分层和分类,构建路线图,以及对路线图的维护和更新。这种工具的进一步发展应该是能够自动按照文档和人员的变化,增删或者移动知识路线图的内容。

4) 知识的验证与评价

不论是从外面采集的知识,还是内部生成的新知识,都需要进行验证与评价。最根本的验证与评价要通过应用知识的实践,但在应用之前,为了做出是否应用的决策,还得通过对知识有发言权的人(包括专家和使用者的)检验和评价。系统中的这一模块就是为此而设的。它能够收集专家和使用者的对知识的真实性、可用性以及水平和质量作出的判断和评价,以提供选择时的参考。

5) 知识的“推”送

用户既可以从系统中“拉”出所需的知识,又可以由系统向他“推”送知识。系统中的知识推送模块起到一个订阅中心的作用,它列出所有可供用户选择的知识内容,供用户选订。当有新的资料进入系统后,它还能把更新的信息通知用户。

6) 实时沟通

实时沟通模块提供了用户通过网络进行实时沟通的渠道。它既可以进行个人之间点对点的沟通,也可以通过远程会议的方式,进行工作讨论和会商。由于现在网络的频带日益加宽,所以通过摄像头和话筒等工具,可以形象地传递双方的沟通内容乃至会场气氛和人的情绪,强化了会议的真实程度。这对于进行隐性知识交流和共识的形成是一种极其有效的工具。

7) 团队工作管理

由于现在组织中的知识创新工作多半是由一些创新团队担任的,知识团队都是按任务临时组织,任务完成后就解散的;另一方面,在组织的其他经营活动中也采取项目管理的方式,项目组也带有临时特点。这就提出了新的问题:怎样能够迅速组建团队,迅速获取有关的知识,而在团队解散后又能把工作过程与工作成果作为组织的知识资源加以保存和保护呢?这就需要对团队的工作进行管理和跟踪。团队工作模块具有自行积累

经验的能力，会向整个系统不断提供经验型的知识。

8) 网上培训

网上培训也是系统所能提供的为更广大的人群服务的模块。由于当前知识更新的速度越来越快，组织培训的任务也越来越迫切和越来越重要。利用网络进行培训是一种成本低而灵活性高的方式。网上培训可以有两种方式：一种是有计划、有进度，对接受培训者要求较严，而且有规定的作业和考核；另一种是由接受培训者自己掌握内容和进度，主动接受测验与考核。两种方式都应该有在线的讨论，只是前一种组织的比较严密而后一种比较松散。前一种方式适合企业有计划地组织培训以提高员工的能力与水平，后一种则更多地带有深造性质，更尊重员工的自愿原则。

除了上面列举的几种应用之外，还有其他一些针对某一特定要求的模块。知识运作与管理的信息技术系统目前也还是在发展的初期，随着其在企业中的地位和作用日益提高，还会有更多的应用出现。

3. 功能层

功能层介于应用层和数据层中间，它不直接参加应用操作，没有用户接口，而只是为用户使用和开发应用任务提供平台和工具支持。这一层包括下列四种功能。

(1) 用户管理。通过用户目录，存储用户的基本信息，如用户的组织归属、个人角色、信息与知识权限等，都在这里登记保存。一旦用户进行登录，就可以统一进行身份验证。对用户进行管理，还有利于组建和扩充社会网络。

(2) 知识流程管理。这一模块支持用户自己定义各种知识流程，确定文档流转，以及各种传阅机制，还可以将阅读意见附加到文档上。在权限许可的条件下，还能查看其他相关意见文档。它还具有各种推送机制。

(3) 协作管理。这一模块负责对于进行协作的团组或者临时组成的协同工作进行登记，检查身份和权限。还可以把协作历史进行统计，提供协同工作的历史记录，为今后进一步的组织提供参考。

(4) 存储管理。这一模块是为应用层和数据层之间的数据传递服务的。它可以进行数据格式的转换，以及提供各种数据视角。

4. 数据层

数据层是系统的最低层。在这里存储组织所有可编码的知识资源。这些资源可以分成两大类型，一类是狭义的数据，另一类是文档。前一类是结构化的数据，用通常使用的关系型数据库来保存。后一类是半结构化的数据，用原始文档形式保存，如网页就是例子。

在传统的数据库中，数据分割成离散的数据段，而文档的保存却是以文档作为基本

单位。文档可能很长,不具备固定的结构,对于使用文档的用户来说,不必严格按照规定的结构进行输入,但文档的处理却遇到很多困难。以网页为例,人们很难对网页中的某一个或一组信息元素进行处理。现在开始使用的 XML 语言部分地解决了这个问题,为信息处理和检索提供了方便。

在系统的具体实现方面,目前大多系统架构采用浏览器—万维网—数据库(B-W-D)模式。在网络方面,处于一地的组织,多半采用局域网互连的方式。对分散在各地的大型企业集团或机构来说,还得借助广域网进行沟通。

在网络构建时,要考虑开放性、灵活性与可扩展性、实用性。

现在,一个新的概念“云计算”(Cloud Computing)开始引人注目,这个概念就是把整个互联网看作是一台可供用户使用的计算机,向用户提供一切互联网服务和各种信息与知识处理能力。相信在不久的将来就会在其中形成“知识云层”。

12.5 知识流程的外包

所谓知识流程外包(KPO),就是企业把自己的知识项目中的或者其他业务中的一部分知识流程,外包给一些专门从事知识外包业务的组织,由他们承担流程中的知识工作。这些流程往往是需要进行高级的知识收集和分析工作,而这又是本企业难以完成的。从事知识外包业务的组织所做的工作可以说是“知识服务”,可以算作一种新的行业。

当前全球业务竞争日益加剧,产品和服务引入的周期时间越来越短,顾客对提供的产品和服务的质量越来越苛求。知识流程外包服务可以向企业提供有关市场、竞争情况、产品和服务的研究成果,因而使企业缩短了从设计到市场的导入时间;提升了组织在业务管理的有效性;帮助企业快速处理突发性事件。

知识流程外包和信息技术外包、软件外包一样虽然都是业务流程外包的一个专业领域,但知识流程外包所需要的专业知识水平却更高。知识流程外包与知识咨询活动不同的地方在于知识咨询常常是一次性的,时间较短,而知识流程外包是直接参加到企业的业务流程中去,历时较长,有的是连续性、经常性的。

下面是五个典型外包的应用领域。

1) 知识产权研究

知识产权资产管理、技术领域知识产权的前景规划、知识产权授权使用、知识产权摘要和知识产权商业化服务等都是可以外包的知识产权研究工作。另外,像申请专利权是一项专业性很强的工作,如果由申请者自己去做,不但投入的资源比外包多,而且质量也比外包差。这些服务不仅能适用于专利,也适用于商标、著作权和其他知识产权。

2) 分析和数据挖掘服务

企业在经营管理过程中需要的分析和数据挖掘服务, 要求算法设计、统计分析和计算机辅助模拟的应用。如果企业自己没有专门的部门和人才, 进行这样的工作其投入要比外包更多而且效果还不一定好。

3) 金融行业内部业务

在金融行业内部, 知识流程外包已经被用于信用评级、损失保护计算和欺骗分析等诸多方面, 也会应用到对某一上市公司股票的研究。

4) 医药和生物科技

在医药和生物科技行业, 知识流程外包可以用于合同研发外包, 临床试验等。

5) 第三方质量监控服务

依托公平、公正的服务理念, 第三方质量监控服务可以为委托方提供从定位分析、确立体系、问题诊断、问题反馈到改进方案及执行提升的一整套服务, 帮助委托方实现“流程可控、反馈及时、质量提升、客户满意”的最终目的。

一般来说, 提供外包业务的一方就是知识服务机构。服务人员利用书籍、数据库、专家、新闻、电话等多种途径来获取信息, 并对信息进行即时、综合的分析研究, 提升为知识, 最终将报告呈现给客户, 作为决策的借鉴。

对于像中国这样的具有一定智力资源的国家来说, 承接知识流程外包对经济发展有下面的积极作用:

(1) 有利于提升产业结构。知识流程外包服务产业是现代高端服务业的重要组成部分, 具有信息技术承载度高、附加值大、资源消耗低等特点。承接知识外包服务, 可以增大服务业在国内生产总值的比重, 提升产业结构, 节省能源消耗, 减少环境污染。承接服务外包对服务业发展和产业结构调整具有重要的推动作用, 能够创造条件促进以制造业为主的经济向服务经济升级。

(2) 有利于转变对外贸易增长方式。承接对外的知识流程外包服务, 可以扩大服务贸易的出口收入。因为它对资源成本依赖程度比较低, 国外设限不强, 具有快速增长的余地, 从而有望成为出口新的增长动力。

(3) 有利于提高大学生的就业率。服务业吸收劳动力就业占社会劳动力的比重正在逐年提高, 而服务外包作为现代服务业的推动器, 将创造大量的就业岗位, 特别是知识服务外包, 可以缓解知识分子的就业压力。它还将带动政府、高校、企业加强人才培养, 提升劳动力素质, 培养一批精通外语, 掌握世界前沿科学技术, 并且与海外市场联系广泛的人才。

知识流程外包服务需要人才的素质较高, 同时还需要强有力的信息技术装备和充足的数据。

12.6 作为隐性系统的学习型组织

除了以信息基础设施为依托的显性知识系统外,还有一类隐性的、无形的知识系统,那就是企业或其他组织机构中的学习型组织。最近几年来,人们非常注意这类组织的创建,并认为它是提升核心竞争力的有效工具。从系统的观点看来,这类组织乃是一类隐形的知识系统。

12.6.1 组织学习的含义与内容

知识与学习是紧密联系的两个概念,人们获取知识大都是通过学习。不但个人要学习,组织也要学习。特别是当今世界的变化日新月异,不学习就会跟不上时代的步伐,不但个人如此,组织也是如此。学习不仅是获取知识,而且要转变人和组织的观念,重新认识世界,采取适当的行动,以求得生存和发展。对于个人的学习,每个人都有很深的认识和体验,但什么是组织的学习,却并不是人人都了解的,我们在本书的 10.8 节介绍过组织学习的概念。现在再从组织学习的重要性和组织方式谈起。

20 世纪 80 年代以来,各种类型的企业组织都面临优胜劣汰的严峻考验。产品生命周期日益缩短,全球范围内的竞争日趋激烈,使得全球性企业的寿命也在不断缩短。发达国家的中小企业平均寿命不足 7 年,大型企业也不足 40 年。企业重组与并购不断出现,如何才能使企业永葆青春,活力长盛不衰,关键在于能否不断学习、创新,以适应变化。

学习是吸收知识、获得信息的过程。国外有的学者认为:组织的学习就是任何组织单位获取被它认为有潜在应用价值的知识;有的认为组织学习是发现和纠正错误的过程;还有的认为学习是进行能力训练;哈佛大学的爱德蒙德森分析和综合各种不同领域对组织学习的定义,认为组织学习是一个过程,在这个过程中,组织的成员积极主动地应用资料(即与组织相关的信息)来指导组织行为,以提高组织连续适应环境的过程。现在企业所面临的竞争环境是复杂多变的,如何适宜于环境是企业生存的首要任务,但由于观念上的惯性,使得这种适应变得非常困难。

组织学习的内容包括在组织传播知识,向外部学习,从过去经验中学习,系统地解决问题和进行实验五个方面。

1) 在组织内传播知识

由于某些知识需要在组织的成员之间共享,才能使知识发挥更大的作用,这就需要这些知识在组织内部迅速而流畅地传播。学习是一种很有效的知识传播方式,不但能使知识在各成员之间传递,而且在学习过程中经过讨论能够深化各自的理解,并达成共识。

2) 向外部学习

组织的外部有着更丰富的知识,向外部学习可以弥补自己组织知识的不足。向别人学习包括向自己的供应商、自己的客户、自己的合作伙伴、甚至自己的竞争对手学习。中国云南白药创可贴的问世就是一个例子。云南白药是治伤的特效药,但药粉敷在创口上很容易脱落,使患者感到不便。国外的创可贴由于容易贴在创口上因而受到患者的欢迎,因而抢占了云南白药的部分市场份额。后来云南白药的员工发现有的患者在撒上云南白药粉后,用创可贴加以固定,于是从这里受到启发,继而开发出云南白药创可贴来。这就是向用户学到了粘贴方法,向竞争对手的创可贴产品学到了粘贴技术。其实有时候对一些相距较远的事物,也可能得到一些启迪,正所谓“三人行必有我师”。

学习的方式是很多的,如参观展览、参加观摩交流和研讨会等。现在流行的所谓“标杆战略”,其内容就是不断发现、分析、采纳并实施业界最佳管理实践的一种调查研究和学习活动。

3) 从过去经验中学习

一个组织在过去的实践活动中,总是有一些成功的经验和失败的教训,这些经验和教训就存在于组织内部,通过组织学习(也就是通过总结)把它们提炼出来,使成功的经验得以继续运用和发展,但注意不要形成“无意义的成功”。对失败的教训要引以为戒,通过“有价值的失败”的思维模式学习,使得今后不要再犯同样或者类似的错误。也就是说,对过去的反思是能够增长自己的智慧的。

4) 系统地解决问题

解决问题包括发现问题、分析问题、寻找方案和最终通过执行使问题得到解决。在这一系列的活动中,人们不断地在通过研究讨论获得经验和知识。所以解决问题本身就是一个组织的学习过程。在这个过程中不但有个人和集体之间的知识转移,而且也有显性知识与隐性知识之间的转化。所谓在干中学,对于组织的学习来说更有实际意义。

5) 进行试验

试验也是一种组织学习方式。由于一种变革或者创新常常涉及的面很大,如果全面推开,风险太大,可以限在一定范围内进行试验。试验不但可以验证原来的预想和设计正确与否,而且也是一种宣传教育和知识传播的工具。

12.6.2 学习型组织的特点

学习型组织的产生是由于经济的全球化,组织所处的环境经常发生急剧变化,组织间的竞争日益激烈而引起的。因为面对这样的环境,需要一种灵活的、适应性强的组织结构。学习型组织由于强调不断学习和创新,比较能适应这种要求。

学习型组织具有下列五种卓越的能力:

- (1) 认知能力;
- (2) 适应能力;
- (3) 创新能力;
- (4) 协作能力;
- (5) 沟通能力。

从学习型组织的结构和工作方式来看,它具有下面四种特点。

(1) 具有明确的愿景和清晰的战略目标。在五项修炼^[4]的第三项中,要求组织具有共同愿景。这种愿景就是组织的战略目标。有了这样的目标,就能使具有不同专业、不同经历,乃至不同个性的人聚集在一起,为了共同的目标而奋力工作。

(2) 具有较强的竞争理念。由于组织和它的成员总是在不断地学习和处于不断进取的状态之下,所以面对竞争环境,有着较强烈的竞争观念,不愿甘居下游。

(3) 在组织结构上,具有适合团队创造性工作的扁平结构或网络结构。这样的结构打破了层次结构或职能结构对创新工作的约束,减少了制度成本,提高了效率和效能。

(4) 具有先进的组织文化。学习型组织中的人与人的关系是平等的,组织中充满了民主气氛,使大家可以毫无拘束地各抒己见,畅所欲言。

在学习型组织中,领导人不是高高在上,而是与下属平等相处,他的作用很像乐队的指挥。乐队中每一种乐器的演奏都不需指挥去具体指导他(演奏家在具体乐器演奏上比指挥可能还更熟练和高超),指挥的任务是协调整个乐曲的进行。

圣吉^[4]对组织学习看得很理想化,几乎达到一种超脱的境界。他认为组织成员应没有任何私心杂念,这对于一般的组织而言是难以实现的。虽然如此,但他的这一套理念仍对现代组织具有积极的参考价值。

在学习型组织的推行过程中,也有一些片面强调形式的做法,如规定学习时间和学习内容等,这是舍本逐末的做法。学习型组织应该领会的倒是其精神实质。

12.6.3 实践社群

有一些非正式组织,在知识的管理中也会发挥不可忽视的作用。这类组织是由一些知识工作者按照共同的兴趣或者共同的观点和见解而形成的。由于这是在实践过程中形成的群体,所以叫做实践社群(COP, Community of Practice),或者叫做实践共同体。

知识共享并不是简单地按照行政命令执行,特别是在一些大的组织和竞争激烈的领域中。完全按照领导的意旨组织团队是很难做到一下子就能协调顺利开展工作的,常常需要一段磨合时间。但是如果一些共同关心某项任务或某个课题的人自愿走在一起,把需要的知识和具备知识的人自然组合起来,大家在一个平等而可以自由交换意见的环境中进行知识交流与共享,并且通过毫无拘束的讨论和交流酝酿新的思想,将是一种非常

有利于创新的组织形式。

在一个组织内部,许多成员都有贡献自己的意见并希望得到重视和意见被采纳、付诸实现后见到效果的愿望。但是传统的等级壁垒和部门壁垒限制了这种愿望的实现。如果领导鼓励这种非正式组织的形成和积极活动,对组织的工作、特别是创新活动是大有好处的,而且也能大大提高员工的积极性和创新的主动性,增加组织的凝聚力。

实践社群既可能是正式的,通过自下而上的串联和自上而下的认可而正式成立,并且有一定的组织结构和管理方式;也可能是非正式的,是一种由感兴趣而热心的人们自愿聚集在一起活动,而受到领导的鼓励和支持的(也可能领导并不知道,但活动已经在不断进行的)。这种形式的组织比较松散,成员随时有进有出。

实践社群的成员不但可以是组织内部成员,也可能有外部人员主动参与,这样也会扩大组织与外单位的联系。

实践社群的人数可多可少,一般最多以 20 人为宜,人太多了就不容易集中意见,会分化为几个社群。社群的活动以讨论为主,有时候也会采取调查、访问的形式或者请人作报告的形式。成员之间的沟通可以是面对面的聚会,这样容易沟通隐性知识,也可以通过互联网,如用电子邮件来互通信息。一些正式一点的社群甚至有自己的网站。

实践社群可以有三种形式。

(1) 产品与技术类社群。这类社群所关注的是企业某一种特定的产品或技术的创新和开发。例如,有关新材料的研究,新药品的开发,新型供应链的研究开发,等等。参加这类社群的人多半是凭自己的兴趣,而不是职务的要求。活动中常常由承担研究开发任务的知识工作者提出要求或问题,而社群中的专家经过思考提出意见,再在社群的集体中讨论或者个别交换意见,加以丰富充实。

(2) 角色或功能类社群。这类社群一般是围绕着某一种角色,如项目经理、培训师而形成。参加者是按照职务的指派而参加或退出的。它的功能是帮助所担任的角色出色地完成使命。

(3) 面向产业或市场的某一部门的社群。这类社群是围绕企业当前所关注的市场中的某一个部分而形成的,成员基本致力于支持这一部门的工作。

还有一种所谓特殊兴趣群组(SIG, Special Interesting Group),即指社群中的分组,专门用于研究某一特别感兴趣的问题。参加者不一定是社群的成员,可以吸收外界人士参与。

由于实践社群在组织的知识系统中处于非常重要、但不是极其显著的地位,所以组织应该大力支持社群的形成、建立和活动。组织要为社群活动提供时间和空间条件,如给以一定的自由活动时间(美国的 3M 公司允许员工用 15%的时间从事自己感兴趣的工作),提供一定的聚会场所甚至实验室,并允许在企业网站上建立交流园地。

除了物质条件的支持和保证外,精神上的支持更为重要,组织应该对实践社群的积极参加者加以表扬,顶住外来的冷嘲热讽,有了成果及时表彰。形成一种尊重知识、尊重知识创新的氛围。

现在国外许多组织非常重视实践社群的组建。例如,世界银行近两三年内就建立了130个实践社群。虽然这些社群是非正式的,但是为了促进它的成长,世界银行还是力求促进它们的正规化。他们的经验表明,实践社群成功的关键在于:

(1) 社群的战略具有可操作性。这就要求社群的活动围绕组织的技术与管理创新和工作改进的任务。社群的活动内容有可能涉及组织的总体目标与战略,也可能仅仅涉及某一具体难题,但无论如何,都得与实践应用相结合。

(2) 组织形式可以多样化,按照实际要求采取灵活的形式。

(3) 技术支持也是保证成功的重要条件。虽然现在的信息工具已经能够为实践社群提供多方面的支撑,但是对于人与人之间深入交流的支持还是困难的。

(4) 合作关系是保证实践社群工作的另一个重要条件。现在每一个社群都在自发地寻找合作者,这样可以扩大知识范围,从中得到更多的启示。

(5) 社群的工作要讲求质量,因此应该有一个评估体系与评估方法。从理论上来说,只要对输入、活动过程、输出、成果和影响进行测量和评价,就能够进行评估,但是要真正实现也比较困难。



思考与讨论题

- (1) 你觉得在什么情况下知识战略会反过来影响企业的总战略?
- (2) 知识管理系统的开发与信息系统的开发有什么相似之处,又有什么原则上的不同?
- (3) 一个理想的知识管理系统开发团队人员的知识结构应该是怎样的才好?
- (4) 试草拟一个研究节能技术的科研团队知识管理系统的开发计划。
- (5) 知识管理流程的外包都有什么优、缺点?有什么风险?
- (6) 你同意把学习型组织看做是一个隐形的系统吗?它有哪些系统特征?



参考文献

- [1] 王众托. 知识系统工程[M]. 北京: 科学出版社, 2004.
- [2] 王众托. 知识管理[M]. 北京: 科学出版社, 2009.
- [3] 奉继承. 知识管理: 理论、技术与经营[M]. 北京: 中国经济出版社, 2006.

[4] 圣吉 (Senge P.). 第五项修炼——学习型组织的艺术与实务[M]. 上海: 上海三联书店, 1998.

[5] Garvin D A. Building a Learning Organization [J]. Harvard Business Review, 1993, 71(4), 78-92.

[6] Edmondson A, Bertrand M. From Organizational Learning to the Learning Organization [J]. Management Learning, 1998, 29(1): 5-20.

[7] Probst G. Managing Knowledge [M]. New York, USA: John Wiley & Sons, Ltd. 2000.

[8] Argyris C. Teaching smart people how to learn [J]. Harvard Business Review, 1991, 69(3): 99-109.

[9] Argyris C. Empowerment: the emperor's new clothes [J]. Harvard Business Review, 1998, 76(3): 98-105.

[10] Huber G. P. Organization Learning: The Contributing Processes and the Literature [J]. Organizational Science, 1991, 2(1): 88-115.

反侵权盗版声明

电子工业出版社依法对本作品享有专有出版权。任何未经权利人书面许可，复制、销售或通过信息网络传播本作品的行为；歪曲、篡改、剽窃本作品的行为，均违反《中华人民共和国著作权法》，其行为人应承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护权利人的合法权益，我社将依法查处和打击侵权盗版的单位和个人。欢迎社会各界人士积极举报侵权盗版行为，本社将奖励举报有功人员，并保证举报人的信息不被泄露。

举报电话：(010) 88254396；(010) 88258888

传 真：(010) 88254397

E-mail: dbqq@phei.com.cn

通信地址：北京市万寿路 173 信箱

电子工业出版社总编办公室

邮 编：100036